A structural and functional study of the second periplasmic loop P2 of MalF in the maltose transporter of *Escherichia coli*.

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

im Fach Biophysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der

Humboldt-Universität zu Berlin

von

Tomas Jacso, Diplom-Ing.

Präsident der Humboldt-Universität
Prof. Dr. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Prof. Dr. Lutz-Helmut Schön

Gutachter: 1. Prof. Bernd Reif
2. Prof. Erwin Schneider
3. Prof. Udo Heinemann

Tag der mündlichen Prüfung: 09.07.2010

Zusammenfassung

ABC (ATP-binding-cassette)-Transporter katalysieren den ATP-abhängigen Transport diverser niedermolekularer Substanzen durch die biologische Zellmembran. Ihr Vorkommen erstreckt sich auf alle drei Domänen des Lebens. Der Maltose Transporter von *E.coli* gehört zu dieser Superfamilie der ABC-Transporter. Die Kristallstrukturen des Transporters $MalFGK_2$ wurden kürzlich gelöst für dessen inaktiven Zustand als auch für dessen katalytischen Zwischenzustand. Um den Transportmechanismus besser verstehen zu können, müssen die Kristallstrukturen des Transporters und seiner Komponenten unter physiologischen Bedingungen genau geprüft werden, um den daraus katalytischen Mechanismus zu bewerten. Im rahmen der Dissertation konnte mittels Lösungs-NMR kann gezeigt werden, dass die periplasmatische Schleife P2 von MalF eine unabhängige Faltung aufweist und eine wohl definierte Tertiärstruktur einnimmt, die vergleichbar ist mit der im Kristall vorliegenden Konformation. MalF-P2 interagiert unabhängig von der Transmembranregion von MalF und MalG mit dem Maltose-Bindeprotein in An- und Abwesenheit des Substrats mit einem K_D im mikromolaren Bereich. NMR Untersuchungen zu den an der Interaktion beteiligten Aminosäuren stehen in Einklang mit den Kristallstrukturdaten. Die Analyse residualer dipolarer Kopplungen (RDC) zeigt, dass die Konformation der zwei individuellen Domänen von MalF-P2 in Abwesenheit von MalE erhalten bleibt und der im Kristall ähnelt. Die Zugabe von MalE induziert eine Änderung der relativen Orientierung der zwei Domänen von MalF-P2 um so dem räumlichen Anspruch des Liganden gerecht zu werden. Besonders betroffen hiervon ist die Domäne 2 von MalF-P2, deren Konformation abweicht von der in der Kristallstruktur. Die Struktur der Domäne 1 dagegen bleibt konserviert, während sich lediglich ihre relative Orientierung zu Domäne 2 ändert. MD Simulationen des MalF-P2-MalE-Komplexes deuten auf eine stark dynamische Interaktion von MalF-P2 mit der MalE Bindungsregion hin. NMR CPMG Kinetikstudien weisen auf die Bildung eines ungewöhnlichen Knicks in α-Helix α2 während der Assoziation hin. Diese konformelle Änderung der α-Helix findet auf einer Zeitskala von Millisekunden statt, was im Einklang mit der Austauschrate der Komplexbildung ist.

Summary

ABC (ATP-binding-cassette)-transporters catalyze the ATP-dependent transport of diverse solutes across the cellular membrane. They are present in all three kingdoms of life. The *E.coli* maltose transporter belongs to the ATP binding cassette (ABC) transporter superfamily. Recently, the crystal structures of the full transporter $MalFGK_2$ in its resting and a catalytic intermediate state was solved. At the present state of research, it is of particular interest to scrutinize the X-ray structures of the transporter and its components under physiological conditions as well as to evaluate their implications for the catalytic mechanism.
In the context of the PhD thesis, it could be shown using solution-state NMR that the periplasmic loop P2 of MalF folds independently in solution and adopts a well-defined tertiary structure, which is similar to the one found in the crystal structure. MalF-P2 interacts with the maltose binding protein, independent of the transmembrane region of MalF and MalG, with a K_D in the μM range, in the presence and absence of substrate. NMR studies showed good agreement of the residues interacting in solution to those identified in the X-ray structure. Analysis of residual dipolar coupling (RDC) experiments shows that the conformation of the two individual domains of MalF-P2 is preserved in the absence of MalE, and resembles the conformation in the X-ray structure. Upon titration of MalE to MalF-P2, the two domains of MalF-P2 change their relative orientation in order to accommodate the ligand. In particular, a conformational change of domain 2 of MalF-P2 is induced, which is distinct to the conformation found in the X-ray structure. Domain 1 retains its structure but changes its relative orientation to domain 2. MD simulations of the MalF-P2 – MalE complex show a highly dynamic interaction of MalF-P2 to the MalE interface. From NMR CPMG kinetic studies, a peculiar kink of α-helix α2 can be seen introduced upon association. The transition time of this conformational change of the α -helix is on the ms timescale, which is matching the exchange rate of the complex formation.

Table of contents

Abbreviations

ABC	ATP-binding cassette
ADP	adenosine diphosphate
ATP	adenosine triphosphate
AUC	analytical ultra centrifugation
BP	binding protein
BSA	bovine serum albumine
CD	circular dichroism
Cm^r	chloramphenicol resistance
CO	carbonyl carbon
COSY	correlation spectroscopy
CPMG	Carr-Purcell Meibloom-Gill
D_2O	deuterated H_2O
ddH_2O	*aqua bidest*
E.coli	Escherichia coli
EPR	electron magnetic resonance
FPLC	fast protein liquid chromatograph
Ig-like	immunoglobuline-like
IPAP	in-phase anti-phase
IPTG	isopropyl-thio-b-D-galactoside
ITC	isothermal titration calorimetry
HetNOE	heteronuclear nuclear Overhauser effect
His_6-tag	six-histidine tag
HMQC	heteronuclear multiple quantum correlation
HSQC	heteronuclear single quantum correlation
kDa	kilo Dalton
MalF-P2	periplasmic loop 2 of MalF
$MalFGK_2$	the maltose transporter of *E.coli*
MBP	maltose binding protein (MalE)
MS	mass spectrometry
NBD	nucleotide binding domain
Ni-NTA	nickel nitrilotriacetic acid
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
OD	optical density
PAGE	polyacrylamide-gelelectrophoresis
PCR	polymerase chain reaction
pdb	protein database
RDC	residual dipolar coupling
rmsd	root-mean square-deviation
RT	room temperature
SDS	sodiumdodecylsulfate
SN	signal-to-noise
τc	correlation time
TM	transmenbrane

TMD	transmembrane domain
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation-optimized spectroscopy
w/wo	with and without
wt	wildtype

Standard chemical or physical abbreviations and symbols are not listed. Amino acids are abbreviated by the one- or three- letter code. Nucleotides are abbreviated by the one- letter code.

1. Introduction

1.1. Biology

This first section will provide a brief introduction to structural studies of membrane proteins and a section containing a general overview of the ATP-Binding Cassette (ABC) transporter superfamily. Following, the maltose uptake system of *E.coli* and the maltose transporter $MalFGK_2$, which was the topic of this PhD thesis, will be described.

1.1.1 Membrane proteins

Membrane proteins are key players in a cell's interaction with its surrounding. They perform a wide array of functions as receptor proteins, transporter channels, electrical and photo-inducers[1-3]. Prokaryotic and eukaryotic genome searches show that roughly 30% of all encoded proteins are in fact membrane proteins[4,5]. Many disorders are directly dependent on mutations of membrane proteins and it is not surprising that over 50% of all administrated drugs target this class of proteins[6-9]. The pharmaceutical industry thus has a great interest in high-resolution structures that provide important information about function and mechanistics of membrane proteins. The first structure of a membrane protein, bacteriorhodopsin, was reported in 1975 by Henderson *et al.*[5]. The structure was produced by electron diffraction and resulted in a modest 7 Å resolution structure that nevertheless revealed a bundle of assembled α-helices. To date the two most common techniques to obtain high-resolution structures are X-ray crystallography and NMR spectroscopy. X-ray crystallography is dependent on single crystals of sufficient size and regularity that are difficult to obtain in a dynamic environment such as lipids or detergent micelles. NMR does not need single crystals but is limited to the size of the system (<50 kDa). Though both these techniques have their limitations, they have made major progress during the last decades that has lead to an explosive increase in high-resolution structures of membrane proteins, visualized in Figure 1.1.1.1.
Despite major advances in both techniques, the number of deposited high-resolution structures of membrane proteins is alarmingly low in comparison to their importance. As of February 2009, only 1.7% of all deposited structures in the RSCB databank are membrane proteins, based on OPM[10,11] and PDBTM[10,11] searches. So far 211 structures of unique proteins could be solved using X-ray crystallography and 33 by using NMR[12,13]. The current advance in membrane protein structure determination is thus largely accomplished by X-ray crystallography. Though X-ray crystallographers have made great improvements in sample preparation membrane protein crystallization is still a tedious task. Moreover, recent reviews show that solution-NMR and also solid state-NMR groups are making remarkable progress in characterization of membrane proteins[14-20]. By the combined efforts of the two techniques the future of structural characterization of membrane proteins looks very promising and could one day catch up with the efforts of soluble proteins.

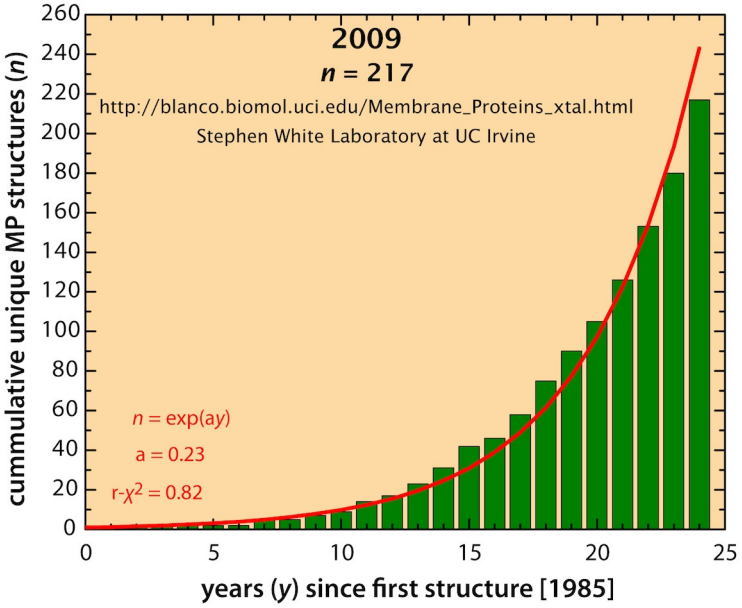

Figure 1.1.1.1. Number of membrane protein structures per year[21].
Only unique structures are included. Figure from the website maintained by the Stephen White lab.

1.1.2 General overview of ATP-Binding Cassette (ABC) transporters

ABC-transporters belong to a superfamily that exists in all three kingdoms of life from prokaryotes to humans[22, 23]. The main function of ABC-transporters is to transport a wide range of substrates over lipid membranes, regulating uptake of nutrients and export of toxins, but they are also involved in non-transport processes as translation, elongation, and DNA repair[24-27]. The ABC-transporter system utilizes the binding and hydrolysis of ATP to translocate the substrates over cellular membranes. The importance of these translocation processes are easily realized from the metabolic costs, which are estimated to be roughly 10 – 60% of the ATP requirements in bacteria[28,29] and humans[30], depending on conditions. ABC transporters are known to play a crucial role in the development of multidrug resistance (MDR), a common clinical problem that occurs in patients suffering from infectious diseases and cancer[31,32]. Prokaryotic and eukaryotic microorganisms involved in multidrug resistance are frequently detected with overexpression of ABC transporters, but inhibition by low-molecular weight compounds of ABC transporters have not yet yielded satisfying results in clinical trials of cancer patients[33,34]. Other human pathologies caused by polymorphism in the ABC genes are cystic fibrosis, hyperinsulinemia and macular dystrophy[31].

1.1.2.1 Characteristics of ABC-transporters

The ABC-transporter system consists of two transmembrane domains (TMDs) of α-helices embedded into the lipid bilayer and two nucleotide-binding domains (NBDs) located in the cytoplasm of the cell. For prokaryotic importers substrate translocation is also dependent on a high-affinity binding protein (BP) that specifically associates

with the ligand in the periplasm for delivery to the appropriate ABC-transporter[35]. The range of substrates is extremely diverse, ranging from small ions to amino acids and peptides[36], and the BPs are highly specific to its substrate or in some cases structurally related substrates, such as maltose and maltodextrins[37]. In gram-positive bacteria, which lack an outer membrane and thus a periplasm, they are either anchored to the outer leaflet as lipoproteins or directly fused to the transmembrane domains of the ABC-transporter[25]. The BPs not only functions as high-affinity binding proteins but can take on other regulatory tasks as in the maltose system of *E.coli* where the substrate bound BP MalE interacts with Tar as a chemoreceptor[38].

Exporters do not require the high-affinity binding proteins, but instead contain large intracellular loops (ICDs) that extend the TM helices roughly 25 Å into the cytoplasm[39]. The intracellular domains keep the two nucleotide binding domains apart in the resting state allowing the substrate to interact with the intracellular parts of the transmembrane domains. The function of substrate recognition for the export systems is far less understood then for its import system counterpart.

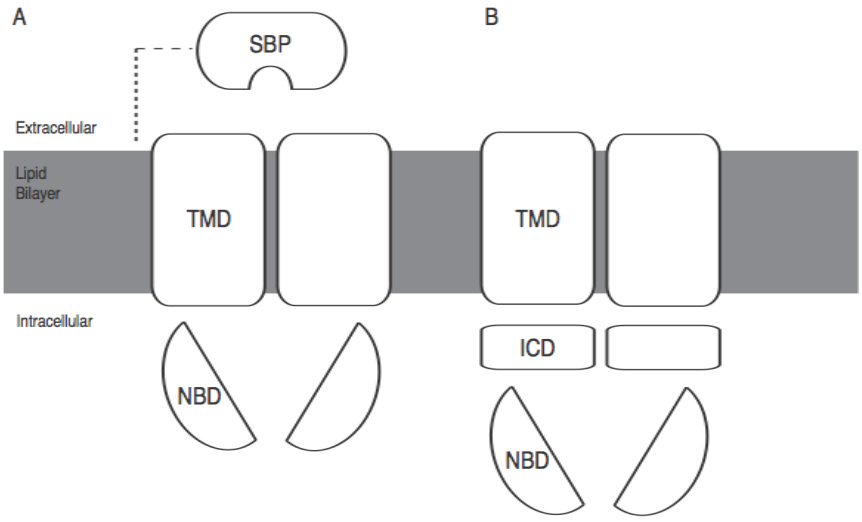

Figure 1.1.2.1.1 The ABC-transporter architecture.
Panel A depicts the ABC importer system with the transmembrane domains (TMDs), the nucleotide binding domains (NBDs) and the substrate binding protein (SBP) for prokaryotic transporters. In gram-negative bacteria the SBP is attached to the outer leaflet of the lipid bilayer as a lipoprotein, depicted as a dashed line in panel A. Panel B depicts the architecture of ABC exporters. Exporters differ from importers in that they have long intracellular loops (ICDs), which extend the TMDs into the cytoplasm.

The sequence and the architecture of the TMDs vary greatly, reflecting the chemical diversity of the translocated subtrates. The number of α-helices for ABC-transporter systems varies between 5 and 10 for each TMD[40]. The α-helix bundle creates the translocation pathway through the lipid membrane and upon ATP-hydrolysis the pathway switches between different conformations allowing access to either the extracellular or intracellular compartments of the cell for the substrate.

In contrast to the great variety in the TMDs the NBDs are highly conserved consisting of a RecA-like and a helical domain, with several characteristic motifs in both domains[40]. These motifs include the Walker "A and B" motifs[41], common to many nucleotide binding proteins, and the LSGGQ-, the stacking aromatic-, H- and Q-loops[39]. These motifs are essential for binding and coordination of ATP for the hydrolysis step where the chemical energy in ATP is transferred into "mechanical" energy for the transporter[42]. The coordination of ATP of the different motifs is shown for the MalK dimer of the maltose transporter in Figure 1.1.2.1.2. Further the two NBD nucleotide-binding sites always face each other and the NBDs are always present as dimers in the fully assembled transporter[25].

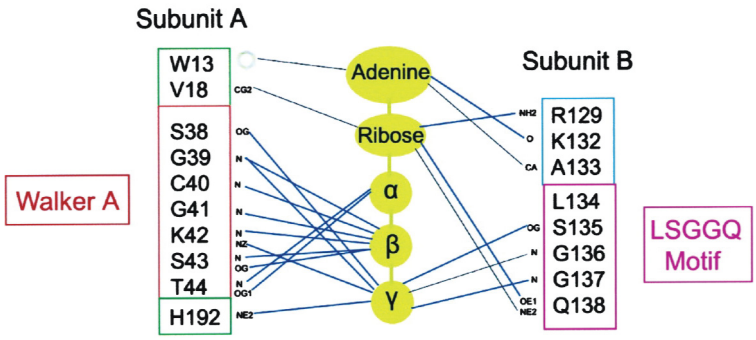

Figure 1.1.2.1.2. Schematic diagram of the interaction between one of the two ATPs bound to the homodimer of MalK in the maltose transporter.
Interactions by the different conserved motifs in the NBD are marked out for their respective subunit, subunit A (RecA-like) and subunit B (helical domain). Black lines represent van der Waals contacts, and blue lines correspond to hydrogen bonds and salt bridges. Figure adapted from Chen *et al.*[42]

The interface between the TMDs and the NBDs of the fully assembled transporters is formed by intracellular conserved coupling helices (EAA-loops), which interact with the grooves on the NBDs' surfaces[43]. This interaction forms a "ball-and-socket" joint that is similar among all ABC transporters although very low sequence homology is apparent. It is therefore believed to be a common mechanism of coupling motions between the NBDs and TMDs that allow large rigid body movements during structural transitions[25,44].
The extracellular loops of the TMDs are also believed to be important for function of ABC-transporters. They can possibly be involved in a wide range of interactions such as substrate recognition, protein complex stabilization and signal transduction from the extracellular parts to the intracellular NBDs. A few examples can be found for the $MalFGK_2$-E transporter. In the crystal structure of a catalytic intermediate of the transporter the periplasmic loops P3 of MalG and P2 of MalF make several important interactions. The loop P3 of MalG is deeply inserted into the substrate-binding pocket of MBP. Oldham *et al.* propose that the P3 loop "scoops" the substrate out by interacting with the binding pocket of MalE[45]. Additionally the insertion of a periplasmic loop into the binding pocket has also been observed in the BtuCD-F crystal structure[46]. In conjunction the loop P2 of MalF interacts with the N-terminal

domain of MalE in a cap-like manner[45]. The specifics of this interaction are not yet understood. Biochemical studies show that the periplasmic loops P2 of MalG and P1 of MalG interact with the substrate-binding protein throughout the catalytic cycle and that the P2 loop also could be responsible for communicating substrate availability to the NBDs[47-49].

ABC-transporters can be structurally characterized by the arrangement of the TMDs. The protein fold can be divided into three categories; the type I ABC importer (e.g. $MalFGK_2$[45] and MetNI[50]), the type II ABC importer (e.g. vitamin B_{12} transporter[51]) and ABC exporters (e.g. Sav1866[52] and MsbA[53]).
Figure 1.1.2.1.3 depicts the crystal structures of $ModB_2C_2$-A, BtuCD-F and Sav1866, which represent the three different folds.

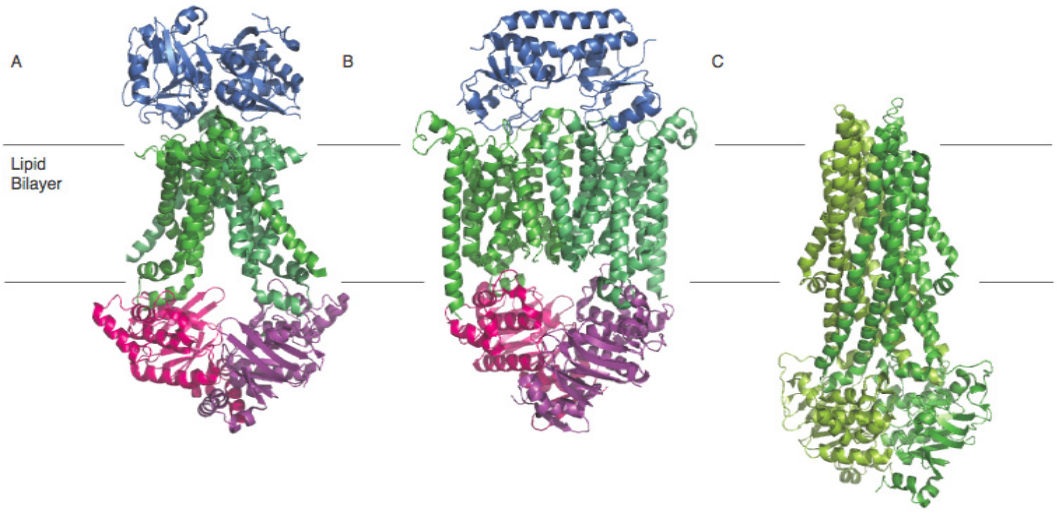

Figure 1.1.2.1.3. Structural characterization of ABC-transporters.
Panels A to C depict crystal structures of $ModB_2C_2$-A[54], BtuCD-F[51] and Sav1866[52], from left to right. The three crystal structures refer to the type I ABC importer, the type II ABC importer and the ABC exporter, respectively. Figure generated with PyMol[55] and pdb codes: 2ONK, 2QI9 and 2HYD, respectively.

In contrast to the classification of ABC-transporters by structure, Davidson *et al.* recently presented a new mode of classification based upon sequence comparison[25]. Three classes can be distinguished; Class I encompasses transporters with the TMD and NBD domains fused together in a single polypeptide chain, Class II includes non-transport ABC-transporters lacking TMDs, and Class III have separate polypeptide chains for the TMD and NBD domains. A comparison of these two different classifications reveals some interesting features that are relevant to the existence of structurally uncharacterized TMD folds[39].

1.1.2.2 Substrate transport

With the available structural and biological data of different ABC systems two different models have been proposed for the translocation mechanism of substrate; the alternating-access model[39] for import systems and the ATP-switch model[56] for export systems. Schematics of the two different transport models are represented in figures 1.1.2.2.1 and 1.1.2.2.2 respectively.

The alternating-access model is based on the ABC importer systems, which include the high-affinity substrate binding proteins and includes three major steps. In the resting state the transmembrane domains of the importer systems are in an inward-facing conformation with the nucleotide binding domains being held open in an outward-facing conformation. The substrate-loaded binding protein docks into the periplasmic side of the transmembrane domains triggering the ATP binding to the nucleotide binding domains and closing their interface. Upon closure of the nucleotide binding domains the transmembrane domains switches to its outward-facing conformation allowing access to the transport pore for the substrate. ATP is then hydrolyzed opening the dimeric interface. Consequently the transmembrane domains switch back into the inward-facing conformation, releasing the substrate and allowing access to the cytoplasm. After release transport of substrate ADP and P_i is released from the nucleotide binding domains, reverting the transporter back into its resting state[39].

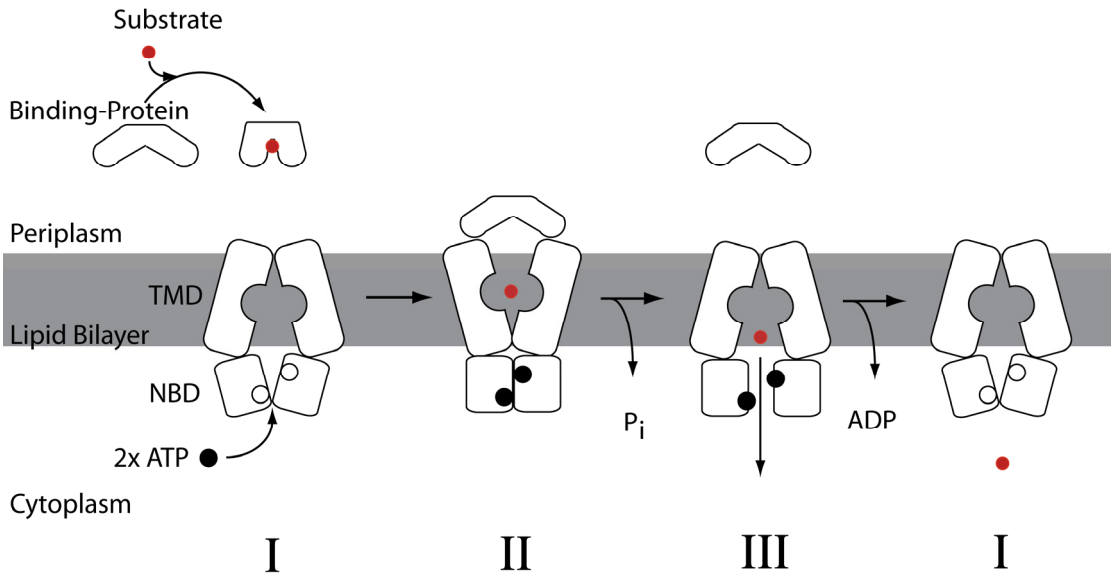

Figure 1.1.2.2.1. The alternating-access model is based on the crystal structures of ModBC-A[54], HI1470/1[57] and $MalFGK_2$[45].
The model contains three major structural changes before reverting back into its resting state.

The second transport model is based on structural and biochemical data of ABC exporter systems. In the ATP-switch model, mainly based on data of MsbA[53], the

transmembrane domains are in an inward-facing conformation in the resting state in consensus with the alternating-access transport model for ABC importer systems. In contrast to the import systems the nucleotide binding domains are relatively far apart to accommodate the amphiphilic and hydrophobic substrates. The intracellular domains of export systems extend the transmembrane domains into the cytoplasm facilitating the space between the nucleotide binding domains contra as seen in the import systems. In MsbA the substrate-binding pocket is large enough to accommodate the relatively large sugar groups of lipopolysaccharides (LPS)[53]. Binding of the substrate and binding of ATP to the nucleotide binding domains, closes the dimer interface, triggering the transport cycle. Upon closure of the dimeric interface of the nucleotide binding domains the transmembrane domains undergoes conformational changes in consensus with the alternating-access model. The sugar head groups are sequestered within the substrate-binding pocket in the transmembrane domains upon the closure of the nucleotide binding domains. The substrate-pocket of MsbA is lined with charged and polar residues energetically favourable for polar moieties in amphiphilic compounds or sugar groups of lipopolysaccharides (LPS) and highly unfavourable for hydrophobic substrates. It is believed that the unfavourable environment for the hydrophobic parts of LPS or substrates causes it to "flip" into an energetically more favourable position with the hydrophobic moieties positioned in the outer membrane leaflet. Upon the conformational changes of the transmembrane domains the rigid-body shearing during the movement could cause the hydrophobic moieties to be dragged through the lipid bilayer. Hydrolysis of one ATP is believed to widen the periplasmic opening of the transmembrane domains and pushing the substrate towards the outer leaflet of the lipid bilayer. Hydrolysis of the second ATP and release of P_i causes opening of the dimeric interface of the nucleotide binding domains restoring the transporter into its resting state with the transmembrane domains in the inward-facing conformation and the nucleotide binding domains separated[53,58].

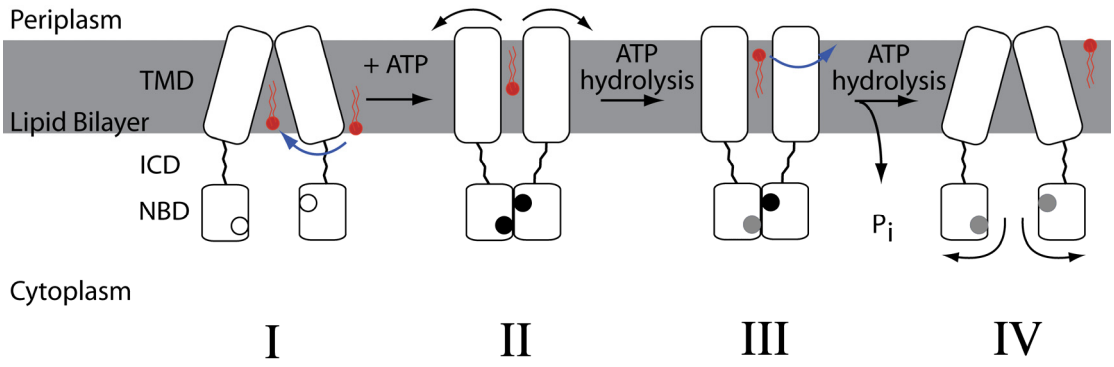

Figure 1.1.2.2.2. The ATP-switch model is based on structural and biochemical studies on MsbA and contains four major structural steps.
Nucleotide-free NBDs are seen as boxes with white-filled circles, ATP-bound NBDs with black-filled circles and ADP-bound NBDs with grey-filled circles.

Complete understanding of how the transport cycle of ABC transporters occur has not yet been elucidated. The key question of how transport of substrates is initialised and if periplasmic or cytosolic parts trigger the catalytic cycle are still one of many questions that needs to be answered. Vast amounts of structural and biochemical data support a model in which ATP binding and hydrolysis is coupled to the conformational changes in the transporter. It is still debated whether substrate binding or nucleotide binding occurs first[25,44,59-62] and if the power-stroke is provided by the closure of the NBDs or by the ATP-hydrolysis[63,64].

Though ABC-transporters are of great interest for the pharmaceutical industry, they have not yet been extensively investigated by structural biology methods. As of date, 17 X-ray structures exists of nine ABC-transporters of unique protein sequences (Table 1.1.2.2.1), none from human[13].

As described above ABC-systems are large multi-protein complexes, and are difficult to access by solution-NMR due to their size. So far, X-ray analysis has been the only method for obtaining structural information. Large uniformly labelled proteins yield spectra of high complexity and thus require tedious analysis, but new NMR based approaches show great promise. To tackle these systems a well suited approach is that of "divide-and-conquer" where the complexity could be greatly reduced by labelling of specific domains and later assembled back for the full system. Intein technology already exists that can provide such labelling schemes[65-67], but now needs to be applied for appropriate complex systems. Another approach is to examine smaller parts of the system within the size limits of solution-NMR. To circumvent the size limit of solution-NMR solid-state NMR could be applied to access these systems. Not only is solid-state NMR not limited by size but it can also acquire dynamic features not accessible by EPR or X-ray. Solid-state NMR provides the possibility to use ABC transporters reconstituted into liposomes or even extracted membranes from the bacteria, greatly simplifying purification protocols and allowing structural studies in native environment. Though Solid-State NMR seems like an ideal technique for these systems technical issues still need to be addressed before it can be used on regular basis for membrane protein systems.

Table 1.1.2.2.1. Currently determined ABC-transporter structures as of Dec 2009.
The structures are presented with their pdb code from the RSCB data bank[70], experimental method and reference. Data from the White lab homepage[13].

System and host	Pdb code	Method	Reference
BtuCD Vitamin B_{12} *Escherichia coli*	1L7V	X-ray	Locher *et al.* 2002[51]
Sav1866 *Staphylococcus aureus*	2HYD	X-ray	Dawson & Locher 2006[52]
ModB_2C_2-A *Archaeoglobus fulgidus*	2ONK	X-ray	Hollenstein *et al.* 2007[54]
HI1470/1 *Haemophilus influenzae*	2NQ2	X-ray	Pinkett *et al.* 2007[57]
MsbA-AMPPNP *Salmonella typhimurium*	3B60	X-ray	Ward *et al.* 2007[53]
P-glycoprotein *Mus musculus*	3G5U	X-ray	Aller *et al.* 2009[68]
MalFGK_2-MBP *Escherichia coli*	2R6G	X-ray	Oldham *et al.* 2007[45]
MetNI *Escherichia coli*	3DHW	X-ray	Kadaba *et al.* 2008[50]
FbpC *Neisseria gonorrhoeae*	3FVQ	X-ray	Newstead *et al.* 2009[69]

1.1.3 The maltose uptake system of *E.coli*

The rod shaped gram-negative bacteria *Escherichia coli* is commonly found in the gastrointestinal system of warm-blooded organisms, where it scavenges its nutrients from the hosts metabolic system. One of these nutrients is maltose or maltodextrins and is regulated by the maltose system of the bacteria. The maltose system has been known for over 40 years and involves a myriad of molecules and regulatory phenomena with only a few being mentioned here[71]. Gram-negative bacteria are characterized by its cell wall. The cell wall contains two membranes, an outermembrane and an inner membrane. The outer membrane is made up of a lipopolysaccharide layer to a peptidoglycan layer, which is often associated with the pathogenic abilities of the bacteria[72]. The inner membrane, which is entirely made up of peptidoglycan, is much thinner than found in gram-positive bacteria, which only have one membrane layer[72]. The two layers are separated by a periplasmic space, which can constitute up to 40% of the total cell volume and has a loose network of murein chains and also contains degradative enzymes[73].
The uptake of maltodextrins of *E.coli* requires multiple steps of transport over the two membrane layers before the bacteria can utilize it in its metabolic pathways. Figure 1.1.3.1 provides a coarse schematic of the elaborate transport systems of the scavenged substrates into the cytoplasm from the cells surroundings.

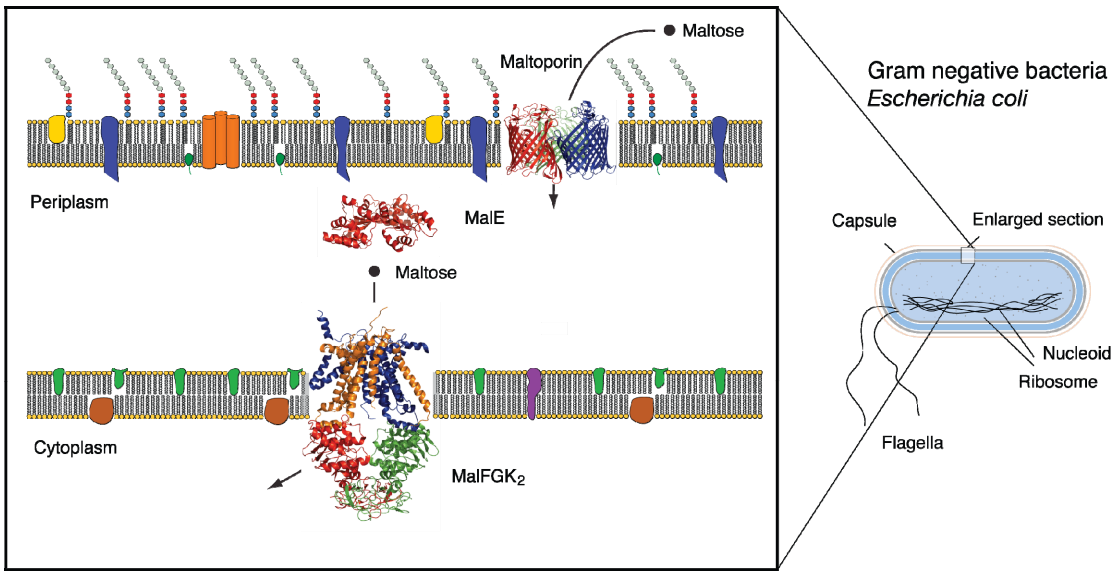

Figure 1.1.3.1. Schematic of the maltose uptake system of *E.col.*
Maltoporin allows maltose/maltodextrins to diffuse through the outer membrane from the bacteria's surroundings. The periplasmic substrate-recognition protein MalE binds substrate at high-affinity and functions as the recognition site for the ABC transporter $MalFGK_2$ in the inner membrane. Pdb codes 1MPM, 1OMP and 3FH6 were used for the structures of Maltoporin, MalE and $MalFGK_2$, respectively.

Efficient uptake of maltose/maltodextrins at low concentration in the living environment of the bacteria requires a highly specific diffusion pore into the periplasmic space. The maltoporin, or LamB, mediates this in the outer membrane.

This trimeric 18-stranded antiparallel β-barrel contains a wide channel in each of the subunits[74] and does not only mediate maltodextrins uptake but also other carbohydrates under starvation conditions[75,76]. Crystal structures of maltoporin complexed with maltose, maltotriose or maltohexaose gives a detailed picture of how the porine functions. The pore's channel entrance is lined with a series of aromatic residues that are arranged in a helical pathway into the pore. These aromatic residues are spaced 6-7 Å apart and make transient bonds to hydrophobic faces of sugar molecules guiding the substrate through the pore. The aromatic helical path is also called the "greasy slide" and is surrounded by hydrophobic residues and are assumed to convey sugar specificity[77,78]. The uptake of maltose/maltodextrins from the periplasmic space into the cytoplasm is controlled by the multicomponent and periplasmic binding-protein dependent ABC high-affinity transport system $MalFGK_2$ – E[79,80]. Substrate recognition by the system is determined primarily by the periplasmic protein MalE that occurs at high concentration, ~1 mM and in a 30 – 50 fold excess over $MalFGK_2$[81], in the periplasmic space. The $MalFGK_2$ system can only transport linear maltodextrins up to six glucose units, maltohexaose, whereas MalE also binds cyclodextrins and *p*-nitrophenyl derivates of maltooligosaccharides[82,83]. This is due to the mode of binding of the substrates to MalE. Substrates that are bound at the reducing end and not within the dextrinyl chain are subjected to transport. For efficient uptake of maltodextrins the periplasm also contains the α-amylase MalS, which cleaves maltodextrins but not maltose. Its preferred product released from larger maltodextrins is maltohexaose, the largest maltodextrins subjected to transport by $MalFGK_2$[84]. It thus has the function to degrade larger maltodextrins that entered the periplasm through the maltoporin system to shorter dextrins that can be transported into the cytosol[85]. The ABC high-affinity transport system $MalFGK_2$, which is responsible for the transport of the substrates of MalE will be described later in more detail. Furthermore, the cytosol contains the three enzymes MalP, MalQ and MalZ, which metabolize incoming maltose and maltodextrins into glucose and glucose-1-phosphate, by combined action, which are used in the glucolysis of the bacteria[37].

The major regulatory unit of the maltose uptake system is the cytosolic protein MalT, which acts as a positive regulator. It is activated by an inducer, maltotriose, whose synthesis is directly dependent on the metabolic state of the cell and stimulates transcription by activating RNA polymerase[86]. Other regulatory systems of the maltose uptake system are the global carbohydrate regulation of the cell, which is effected by the cAMPCAP system[37] and the dephosphorylated form of glucose-specific enzyme IIA of the bacterial phosphotransferase system (PTS) that inhibits transport activity by direct inhibition by binding to the MalK dimer of the MalFGK system located in the inner membrane[37].

Table 1.1.3.1. *Mal* genes of the maltose uptake system with their product and function.
Table adapted from Boos and Shuman[37].

Gene	Gene product and function
*mal*T	MalT: transcriptional activator, essential for transcription of all *mal* genes except the *mal*I/X/Y gene cluster. Binds ATP and maltotriose as inducer.
*mal*E	MalE: periplasmic MBP; binds maltose/maltodextrins with micromolar affinity.
*mal*F	MalF: integral membrane protein of the transporter system. In association with MalG and MalK, it forms the $MalFGK_2$ translocation complex.
*mal*G	MalG: integral membrane protein of the transporter system. In association with MalF and MalK, it forms the $MalFGK_2$ translocation complex.
*mal*K	MalK: transport ATPase, responsible for energization of transport. In association MalF and MalG, it forms the $MalFGK_2$ translocation complex. Target of inducer exclusion by unphosphorylated $EIIA^{GLC}$ of the PTS. In the absence of the inducer, it interacts with MalT to cause repression.
*lam*B	Maltoporin: receptor for λ phage and specific pore for maltodextrins (maltoporin, glycoporin).
*mal*M	Periplasmic protein of unknown function partially associated with the outer membrane. Contains an Ala-Pro linker also found in OmpA.
*mal*P	Maltodextrin phosphorylase. Substrates are maltopentaose and larger maltooligosaccharides. *mal*P mutants still grow on maltose but accumulate large amount of maltodextrins under conditions.
*mal*Q	Amylomaltase. Maltodextrinyltransferase with maltotriose as the smallest substrate. *mal*Q mutants cannot grow on maltose, are sensitive to maltose, and are constitutive for *mal* gene expression.
*mal*S	Periplasmic α-amylase, cleaves preferentially hexaose from the nonreducing end of meltotextrins.
*mal*Z	Maltodextrin glucosidase and γ-cyclodextrinase, cleaves glucose sequentially from the reducing end of maltodextrins. Maltotriose is the smallest substrate. It linearizes γ-cyclodextrins but not α- and β-cyclodextrin.

1.1.4 The maltose transporter $MalFGK_2$-E.

One of the best characterized ABC systems is that of the $MalFGK_2$-E transport system. A vast amount of biochemical and structural data have been gathered on the system for the last four decades and has given it a position as a model system for ABC transporters. This bacterial ABC-import transporter consists of the substrate-binding protein MalE, the transmembrane domain proteins MalF and MalG and the ATPase homodimer MalK.The transporter has been characterized structurally and two crystal structures exists of the maltose transporter in its resting-state[87], in absence of its substrate-binding protein MalE, and in a catalytic-intermediate state with MalE tightly associated to the transmembrane domains MalF and MalG[45]. In conjunction there are also crystal structures available for the ATP-Binding Cassette Dimer MalK

in three different conformations[42]. The crystal structures correspond well to the three different catalytic steps in the alternating-access model.

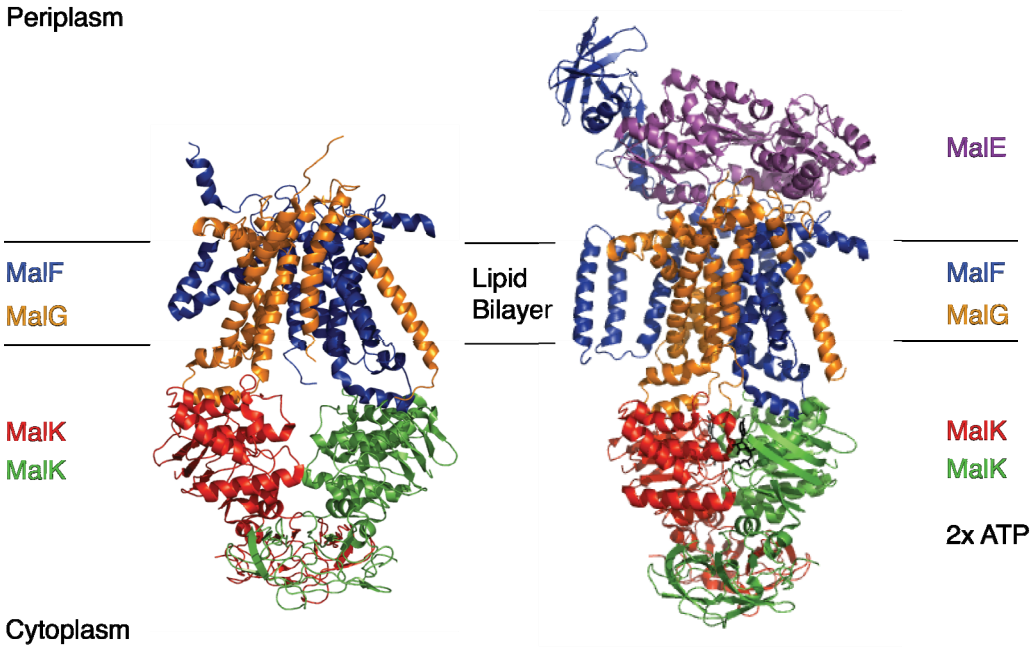

Figure 1.1.4.1. The maltose transporter[45,87].
$MalFGK_2$-E in its resting-state[87] (left) with the TMDs in an inward-facing conformation and with the NBDs open, and a catalytic-intermediate state[45] (right) with the TMDs in an ouward-facing conformation with the SBP tightly associated and the NBDs closed with ATP bound. The maltose substrate is also seen bound in the pore of the transporter, presumably in its binding pocket[45].

The crystal structures of $MalFGK_2$-E show that the conformations of the transmembrane domains and nucleotide-binding domains are in agreement with the postulated transport model of ABC-importer systems, which was based on the crystal structures of ModBC-A[54] and HI1470/1[57].
The transport system utilizes the chemical energy stored in ATP to bring about the conformational changes necessary for transport of substrates. The MalK dimer on the cytoplasmic side of the transporter binds ATP and harnesses this energy into "mechanical" movement of the transmembrane proteins MalF and MalG, alternating them between an inward-facing and an outward-facing conformation. The substrate-binding protein MalE regulates ATP hydrolysis of MalK by interaction with the periplasmic parts of MalF and MalG, triggering the ATP hydrolysis. Studies of the binding protein show that it can interact with the transporter both in its substrate-bound and substrate-free form though both forms have distinctively different conformations[88]. Though both forms can interact with the transporter catalytic activity studies show that maximal rates of ATP-hydrolysis are detected only in the presence of substrate-bound MalE, 1400 and 5 nmol/min*mg for substrate-bound and substrate-free MalE respectively[88]. This old enigmatic question why the transporter undergoes futile transport cycles have not yet been revealed. One possibility is that the futile transport cycle keeps the transcription of the *mal* genes at a basal minimum

at low amounts of maltose and maltodextrins in the cells surroundings. This to keep the maltose uptake system on "stand-by" until more favourable substrate conditions in the cells living environment is available. As earlier described, MalT is a positive regulator of the transcription of the whole maltose uptake system where it interacts with the C-terminal domain of MalK. MalK can shift between two states, one being associated with MalF and MalG at the inner membrane leaflet and one as being associated to MalT in the cytosol[37]. In the presence of substrate MalK is associated with MalF and MalG allowing MalT to function as a *mal* gene activator. In the absence of substrate MalK interacts with MalT and acts as a repressor. A similar mechanism has been seen for the proline utilization system, where PutA, the proline dehydrogenase, functions as a membrane-bound enzyme in the presence of proline and as a cytoplasmic repressor in its absence[89].

1.1.4.1 A transport model for maltose

The MalK dimer sites in a semi-closed state upon binding of ATP and not until substrate-bound MalE binds to MalF and MalG will the ATP-bound MalK dimer close its interface. ATP is found in high concentrations in the cytoplasm, roughly 1 mM[90], and thus it is thought that the MalK dimer resides predominately in its semi-closed form. The fully closed MalK dimer triggers a conformational change of MalF and MalG from its inward-facing conformation to its outward-facing conformation, allowing MalE to tightly associate to the transporter. The substrate now has access to the membrane pore and can bind into the substrate binding cavity of MalF and MalG. After ATP-hydrolysis MalF and MalG switches back into its inward-facing conformation releasing the substrate and shuttling it into the cytoplasm. ADP and free P_i is then released and the transporter goes back into its resting-state, ready for another transport cycle. The postulated three-step mechanism is shown in figure 1.1.4.1.1.

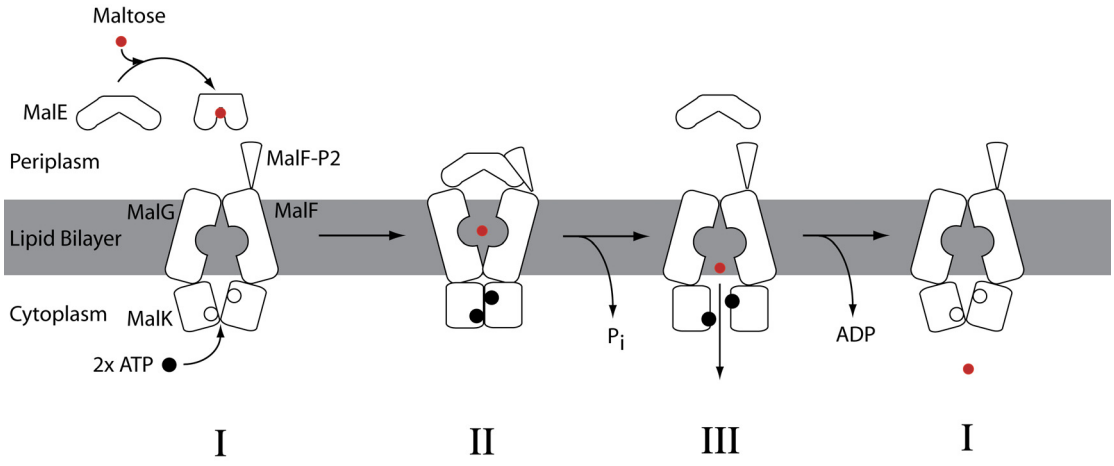

Figure 1.1.4.1.1. The alternating-access transport model adapted for $MalFGK_2$-E.
The three step model show the structural changes of the transmembrane proteins MalF and MalG, switching between inward-facing and outward-facing conformations during the catalytic cycle (steps I – I I I). The figure shows the MalK dimer interface closing upon binding of ATP and substrate-bound MalE (step II), allowing access of the substrate to the membrane pore, and the opening of the dimer interface upon hydrolysis of ATP (steps I I– I I I), releasing substrate into the cytoplasm. After

release of ADP and P_i the transporter goes back into its resting-state (I). Figure adapted from the alternating-access model based on ModBC-A[54] and HI1470/1[57].

1.1.4.2 ATP-hydrolysis

The most pronounced feature of the ABC-transporter system is its ability to harness the chemical energy in ATP into mechanical movements of the transporters. This feature is directly attributed to the nucleotide binding domains and their conserved motifs. The mechanism of hydrolyzing ATP enables the NBDs to be considered as ATPase enzymes. The nucleotide binding domain MalK in $MalFGK_2$-E contain an unique characteristic in the regulatory C-terminal domains towards most other NBDs. This feature stabilizes the dimer interface and Chen *et al.*[42] were successful in utilizing this to crystallize the dimer in three different conformational steps, open, semi-open and closed[42]. The crystal structures show the closure and opening of the dimer as a "tweezers-like" motion, with the C-terminal domain as the handle and the N-terminal domains as the "tips" of the tweezers[42]. The switches between the different conformational steps are not a single action motion but instead as a multiple action motion where the closure of the "tweezer tips" is accompanied by an intra-monomer rotation of the helical domains towards the RecA-like domains, enabling the LSGGQ-motif to contact ATP across the dimer interface and fully closing it. Furthermore, a second rotation adjusts the entire NBDs orientation to the regulatory domains. Figure 1.1.4.2.1 show the orientational differences of the NBDs in the closed, semi-open and open-state.

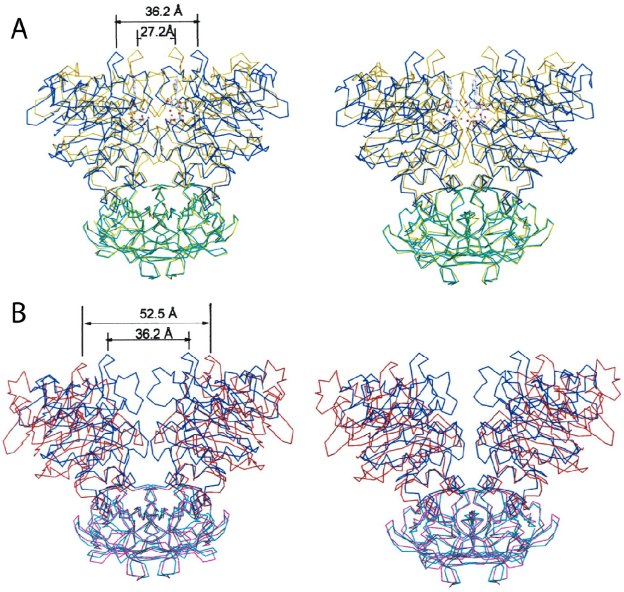

Figure 1.1.4.2.1. Closed, semi-open, and open structures of MalK homodimer with superimposed regulatory domains[42].
The distances between two H89 residues in a homodimer are indicated. **A** show the superimposed closed form with ATP bound, yellow, and the semi-open form without ATP bound, blue. The excellent overlap of the regulatory domains is evident by the green color resulting from the combination of blue and yellow colors. **B** show the overlay of the semi-open, blue, and the open, red, nucleotide-free states. Figure from Chen *et al.*[42]

The conformational changes of the NBDs are transferred to the TMDs through the coupling-helix motif in the transdomain proteins. These helices create the interface between the MalK dimer and the transdomain MalF/G proteins. The coupling helices are situated in the last cytoplasmic loops of the transmembrane proteins MalF and MalG[52]. They are architecturally conserved through the ABC-transporters with a consensus sequence of EAA-X_3-G-X_9-I-X-LP, and are also known as the "EAA"- or L-loops[91]. In the crystal structure of $MalFGK_2$ these two short helices docks into a surface cleft in each MalK subunit and rotates about 30^0 during the inward- and outward-facing conformations[87]. The MalK cleft consists of two helices from the helical subdomain, the helix following the Walker A motif and residues in the Q-loop[42]. The closure of the MalK dimer triggers the reorientation of the coupling helices to their transmembrane helices and rearranging the TMDs from the inward- and outward-facing conformations. The MalK dimer thus produces the "power-stroke" to enable the mechanical movements of the transporter. The docked coupling-helix in the MalK subunit is seen in Figure 1.1.4.2.2 as well as a schematic of the "power-stroke" through the NBDs to the TMDs via the ball and socket joint.

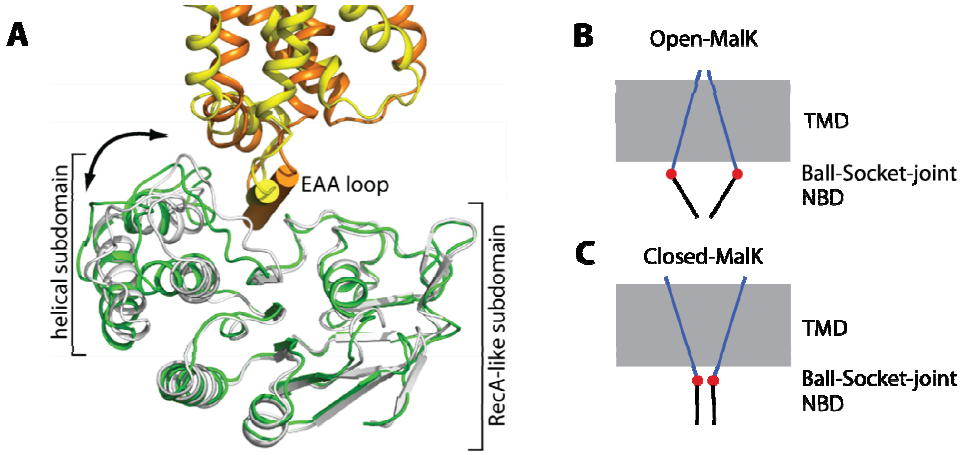

Figure 1.1.4.2.2. The "ball and socket joint" between the transmembrane proteins MalF and MalG and the nucleotide binding proteins $MalK_2$[87].

A The open- and closed-state structures are superpositioned based on the the RecA-like subdomains of MalK. The NBDs of MalK from the inward- and outward-facing structures are shown in green and gray, respectively. MalG is rendered in yellow (resting state) and orange (transition state). The coupling helix is labelled as "EAA loop". **B-C** Show the general principle of the "ball and socket joint" (red filled circle) of the coupling helices between the open- (**B**) and closed-state (**C**) of the MalK dimer (black lines) and MalF and MalG (blue lines). Conformational changes upon closure of MalK are exaggerated in the figure. Figure adapted from Khare *et al.*[87]

1.1.4.3 Characteristics of the transmembrane domains

With the availability of the crystal structures of the maltose transporter, in its resting- and an intermediate-state[87,45], the rearrangement of the transmembrane proteins MalG and MalF can be evaluated in detail. Davidson *et al.* divides each TM subunit into two regions; a core region consisting of the central four helices (TM helices 2-5 in MalG and 4-7 in MalF) and a peripheral region containing helices surrounding the core[87]. The subunits are arranged in an intertwined configuration with a pseudo-2-fold

symmetry, where the peripheral helices of each subunit pack against the other subunits core helices. During the conformational changes the core regions maintain their structures and are seen moving as rigid bodies. The two rigid bodies change their rotation to another by 22^0 with an almost orthogonal rotation axis 45^0 to the membrane plane. The rotation is also accompanied by an 4Å translation along the rotational axes, when using the MalK regulatory domains as a frame of reference[45]. The crystal structures also show a gating mechanism for the substrate to its binding cavity in the pore[45]. Four loops located at the bend of a kinked TM helix creates a hydrophobic gate for the substrate in the inward-facing conformation[87]. A similar gating architecture was also observed in the $ModB_2C_2A$ transporter[54]. The core regions and the gating loops in the inward- and outward–facing conformations are shown in Figure 1.1.4.3.1.

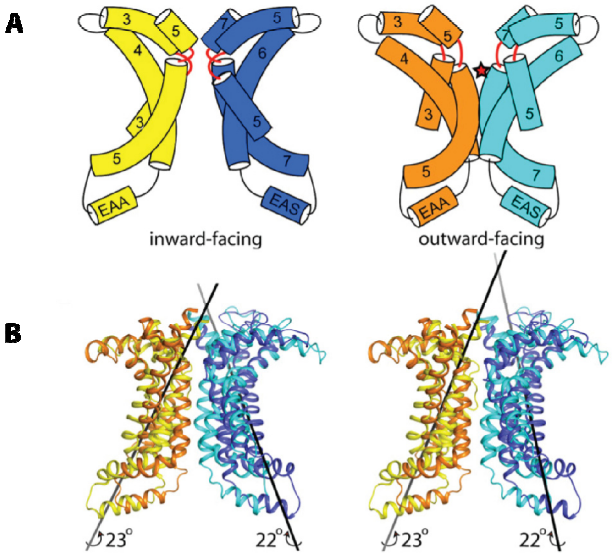

Figure 1.1.4.3.1. Conformational changes in the transmembrane subunits in the catalytic cycle[87].
A, shows the cartoon of the translocation pathway in the resting state (left) and the transition states (right). **A,** red star indicates the location of maltose in the outward-facing conformation. Gating loops are shown in red. The two coupling helices, "EAA"-loops, of MalF and MalG, are labelled "EAA" and "EAS" respectively, based on their sequence. **B**, stereoview of the TM cores: the inward- and outward-facing structures are superimposed based on the MalK regulatory domains. The gray lines indicate the two rotation axes relative to the regulatory domains. Color codes are as follows: MalG resting state, yellow; transition state, orange; MalF resting state, blue; transition state, cyan. Figure adapted from Khare *et al.*[87]

1.1.4.4 The periplasmic region of $MalFGK_2$

Through the crystal structures of the maltose transporter much can be learned about the transport mechanism. Some of the information is however lost, this entails the dynamics and the mobile regions of the transporter. These mobile regions encompass the periplasmic loops of MalF and MalG, which either show very poor or no electron diffraction in the crystal structures[45,87]. These regions have been shown to be important for maltose transport through its interactions with the substrate binding protein MalE[48]. MalE consists of two symmetrical lobes, named the N- and C-lobe

according to their termini[25]. A hinge region between the two lobes allows the two domains to open and close. Upon binding of substrate, which is located in the interior between the two domains, MalE closes making the binding cleft inaccessible to the aqueous surroundings[92].

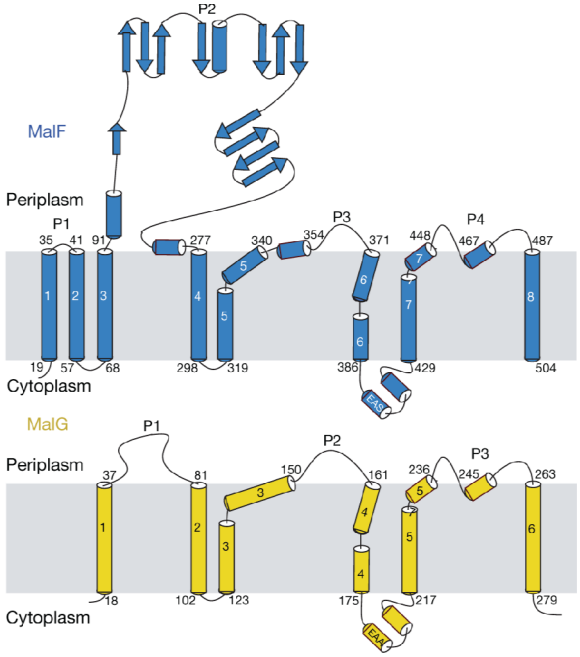

Figure 1.1.4.4.1. Architecture of the transmembrane subdomains MalF and MalG according to the crystal structure[45].
Schematic representation of the secondary structure elements of MalF (blue) and MalG (yellow). The gating loops are located in the kink of helix 5 of respective membrane spanning protein. The first and last residues of each TM helix, predicted to be buried by the lipid acyl chains, are marked out in the figure. Figure adapted from Oldham *et al.*[45]

Recent studies show that the periplasmic loops P1 and P3 of MalG and P2 of MalF are important for substrate recognition and stabilization of the transporter complex and also possibly having other functionalities[47-49]. They were analyzed by tryptic digestions experiments, fluorescence measurements and site-directed cross-linking and was observed undergoing conformational changes and/or large rigid body motions during the catalytic cycle[47-49]. The P1- and P2-loop of MalG and MalF, respectively, binds to one of the two lobes of MalE and their interaction are thought to stabilize the associated complex during the transport cycle[47-49]. As earlier described in section 1.1.2, the P3 loop of MalG is deeply inserted into biding pocket of MalE and Davidson *et al.* postulates that it "scoops" the substrate out from the pocket into the transmembrane pore[45]. The 20 kDa large P2 loop of MalF is an exception among ABC-transporters and is only found in gram-negative *enterobacterial* bacteria. It is seen binding the N-lobe of MalE in a cap-like manner with an Ig-like structure in the catalytic intermediate state crystal structure of $MalFGK_2$[45]. In conjunction the P1 loop of MalG is seen binding complementary to the C-lobe of MalE[47,48]. No data is currently available on the structure or position of the P2-loop in the resting-state or its function during the catalytic cycle. The basis of this work was to characterize the P2-loop's interactions to MalE during the catalytic cycle and to further understand its functional contributions in the transport of maltose through the plasma membrane.

Figure 1.1.4.4.1 show the topology and the location of the periplasmic-, "EAA-", and the gating-loops.

1.2 Objectives of this research

ATP-binding cassette (ABC) transporters are ubiquitous membrane protein complexes that use the energy generated from ATP hydrolysis to transport solutes across the membrane. In bacteria, the majority of ABC transporters are importers that consist of a substrate-binding protein, two integral membrane components and two membrane-associated ATP-binding cassettes. One classical representative is the maltose transporter of *Eschericha coli* / *Salmonella* composed of the periplasmic maltose binding protein (MBP) MalE, the two integral cytoplasmic membrane proteins MalF and MalG and two copies of the ATPase subunit MalK. In *bacteria* and *archaea*, binding proteins are the main determinants of substrate specificity. The determination of the crystal structure of the reconstituted maltose transport $MalFGK_2$ in complex with MalE has recently drawn a lot of attention to this system. There, the large periplasmic loop P2 of MalF (MalF-P2) is in contact with MalE from which maltose has already been released. In fact, MalF-P2 seems to act as a receptor, which recruits MalE and thus maltose, to the core of the membrane protein. An alternative docking site involves the periplasmic loop P3 of MalG, which interacts with the maltose binding site of MalE in the X-ray structure. The question how much MalF-P2 contributes to substrate recognition is still unresolved. Earlier genetic and biochemical investigations show that both MalF and MalG are involved in binding to the substrate-binding protein MalE. It could be demonstrated that the N-terminal lobe of MalE interacts with MalG, whereas the C-terminal part of MalE is in close proximity to MalF.

The present work focuses on the role of the periplasmic loop P2 of MalF. The function of this sequence is largely unknown and structural data of this region is either poor or non-existent in the different catalytic states of the transporter. Yet, there are data indicating conformational changes in this domain for different functional states of the transporter. Based on these data, it was decided to examine MalF-P2 in more detail. This work comprised a study performed on the separately purified, soluble MalF-P2. Data was acquired with a range of analytical and computational methods. A more thorough analysis of the structure was performed with NMR and compared to the X-ray structure of this domain in the assembled transporter. Moreover, the soluble protein was tested for possible interaction partners; especially the interaction with MalE was investigated. The interaction between MalF-P2 and MalE was followed by ITC and NMR to determine the K_D and the timescale of the interaction.

It should be noted that at the beginning of this work, the crystal structure of the full maltose transporter of the catalytic intermediate and resting state was not known. Naturally, its publications in the last year shed new light on structural details of $MalFGK_2$, e.g. with respect to the TMDs' position, contact sites for MalE or possible functions of loop regions. This made it possible to integrate them into a more detailed model of the transport process.

2 Results

This section provides two parts that first describes the expression and purification of the proteins MalF-P2 and MalE together with biophysical methods analysing their respective properties in solution. The second part describes the interaction between MalF-P2 and MalE as seen by NMR, ITC, crystal structure analysis and molecular dynamics simulations.

2.1 Expression and purification of MalF-P2 and MalE

The two polypeptides MalE and MalF-P2 were expressed and purified for this work. MalF-P2 was prepared with different uniformal labelling schemes; ^{15}N, $^{15}N^{13}C$ and $^{2}H^{15}N^{13}C$. Only unlabeled MalE was prepared.

2.1.1 Expression and purification of MalF-P2

The fragment of *E.coli* MalF encompassing basepairs 277-825 (encoding amino acids N93-K275), was amplified by PCR. The fragment covers the entire P2-loop according to a topology model of MalF[93] that was later confirmed by the crystal structure of the complex[45]. It was subcloned into pET15 (Novagen) and verified by sequence analysis. As the NdeI/BamH1 restriction sites of the vector were used, the resulting pMG15 included an N-terminal His_6-tag with a thrombin-cleavage site. The plasmid preparation as well as the purification protocol was provided by Mathias Grote (HU, Berlin).
Upon IPTG-induction of *E.coli* BL21 DE3 <pLysS> an expression of MalF-P2-His_6 can be detected by SDS-PAGE from the cytosol, with an apparent mass of ~20 kDa (Figure 2.1.1.1). Examination for overexpression of the inserted pMG15 plasmid into the *E.coli* BL21 DE3 <pLysS> host cell showed that there is only a small variance of overexpression efficiency after transformation (Figure 2.1.1.1 panel A). Nevertheless, the apparent highest yielding cells were used for up scaling cell cultures.

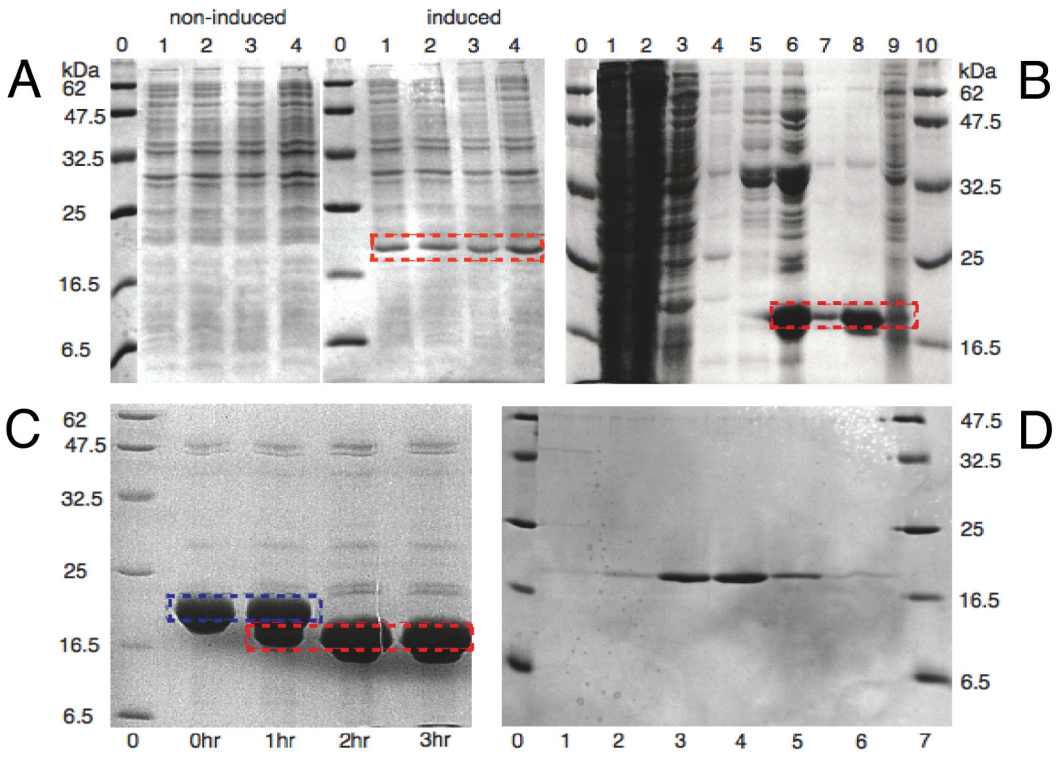

Figure 2.1.1.1. SDS-PAGE overexpression and purification MalF-P2.
Overexpression test of four different cell cultures from four different colonies transformated into BL21 DE3, panel A. Peptide markers are in lane 0 in panels A, C, D with molecular weights marked out in kDa at respective border. Non-induced cell cultures in lanes 1 – 4 to the left and induced cell cultures to the right respectively, after cell lysis. MalF-P2 overexpression is marked with a red dashed box for lanes 1 - 4 of the induced cells. Panel B shows the Ni-NTA-chromatography for MalF-P2. Lanes 1 and 2 show the flow through, lane 3 washing with buffer. Lanes 4 and 5 show the washing with buffer, with 25 mM imidazole added, to prevent unspecific binding. Elution of MalF-P2 with buffer, with 250 mM imidazole added, is shown in lanes 6 – 9. MalF-P2 is marked out with a red dashed box Lane 9 was overloaded for better detection of any residual protein. Thrombin digestion of MalF-P2 at RT is shown in panel C. Lanes 2 - 5 show the digestion of the His_6-tag for 0 – 3 hrs of incubation. Full digestion is completed after 3hrs. Red dashed boxes mark His_6-tag cut MalF-P2 and blue dashed box the uncut MalF-P2. Panel D shows the purity of collected fractions after MalF-P2 was subjected to a Superdex 75 gelfiltration column.

Purification of MalF-P2 via gravity-flow Ni-NTA-chromatography at 5 ^{0}C was successful and showed very low amounts of the target protein in the washing steps and low amount of impurities in the elution step (Figure 2.1.1.1 panel B). Estimated yields for the differently isotopically enriched proteins after purification are listed in Table 2.1.1.1.

Table 2.1.1.1. Protein yields of unlabelled and uniformly labelled MalF-P2.
Yields are in mg/l cell culture after final purification steps.

MalF-P2	unlabelled	^{15}N	$^{15}N/^{2}H$	$^{15}N/^{13}C$	$^{15}N/^{13}C/^{2}H$
yields (mg/l)	21	15	12	14	11

Thrombin cleavage (Amersham Biosciences, Freiburg, D) over night at low temperature 5 ^{0}C, showed 100% cleavage of the His_6-tag and no protein degradation on SDS-PAGE, with an apparent mass of 21 kDa. All purification steps were done at 5 ^{0}C with complete protease inhibitor cocktail tablets (Roche, Mannheim, D), since trace amounts of proteases were always present before gelfiltration.
After successful thrombin cleavage (Figure 2.1.1.1 panel C) the protein was subjected to gelfiltration. The FPLC chromatograms show a single sharp peak of MalF-P2 (Figure 2.1.1.2 and 2.1.1.1 panel D). For unknown reasons a small shoulder on the FPLC chromatogram sometimes could be detected. Likely reason could be overload of the column.

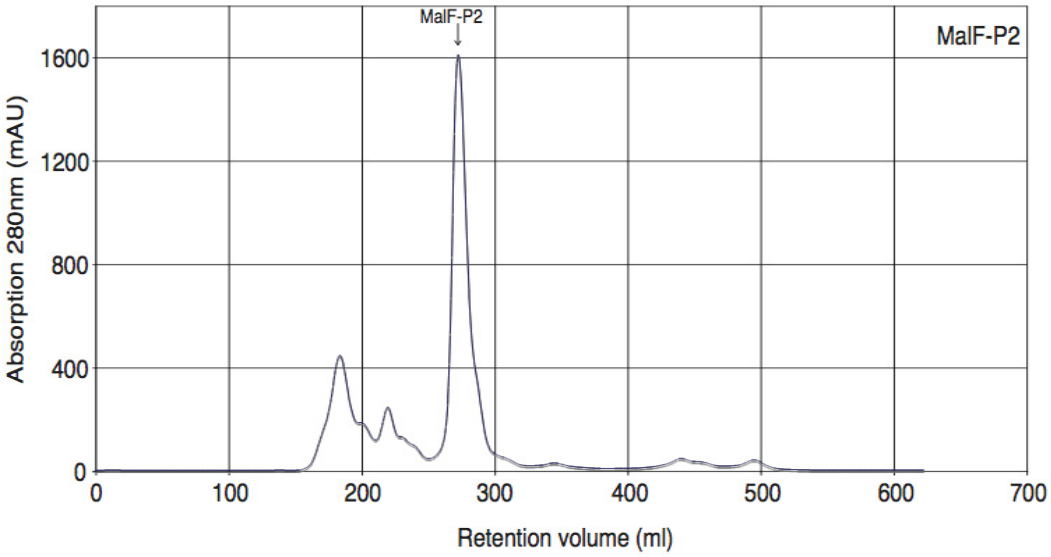

Figure 2.1.1.2. FPLC chromatogram MalF-P2.
Gelfiltration chromatography of MalF-P2 subjected on a Superdex 75 gelfiltration, 500ml column on a FPLC. MalF-P2 comes with a retention volume of 273 ml.

MalF-P2 in 20 mM phosphate 100mM NaCl buffer (pH 7.4) was stable for seven days at room temperature at concentrations ranging between 0.1-1 mM. Protein solutions up to 2mM were stable for two days.
A thin band corresponding to a molecular weight smaller than MalF-P2 was always present in SDS-PAGE to a few percents (1 - 3%), of the major band of MalF-P2, before subjected to gelfiltration. Some preparations showed a small shoulder of the MalF-P2 peak in the chromatogram. Fractions containing this shoulder were discarded. A reason for this could be residual amounts of proteases left in the protein solution.

2.1.2 Expression and purification of MalE

Two constructs were used for overexpression of MalE for this study. Early preparations were carried out as described by Daus *et al.*[94]. These samples were produced by Mathias Grote (HU Berlin). Later on expressions were carried out at the FMP using the plasmid pCB06, which was provided by Anne Diehl (FMP, Berlin)[95].
The MalE-His_6 plasmid pCB06 were inserted into the *E.coli* BL21 DE3 <pLysS> strain and checked for overexpression. Upon IPTG-induction of *E.coli* BL21 DE3 <pLysS> a very high overexpression of MalE-His_6 can be detected from the cytosol by SDS-PAGE, with an apparent mass of ~35 kDa (Figure 2.1.2.1, panel A).
MalE was purified via gravity-flow Ni-NTA-chromatography at 5 ^{0}C, thereby taking advantage of the six histidine residues fused to the N-terminus, Figure 1 panel B. The first column wash showed very low amounts of the target protein and low amount of impurities in the elution step. Yields of around 150 mg/l LB culture were achieved after purification.

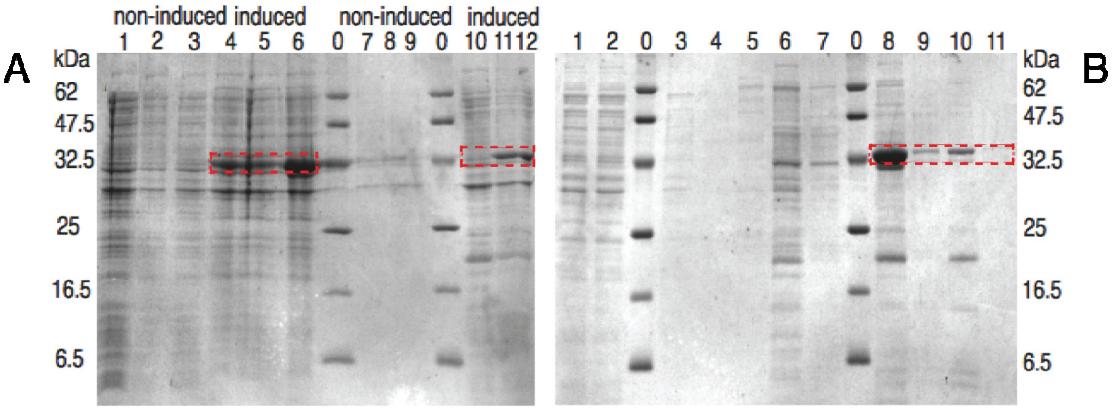

Figure 2.1.2.1. SDS-PAGE overexpression and purification MalE.
Overexpression test directly into three different colonies, A) lane 1 - 6, after transformation into BL21 DE3 cells. Non-induced cells from colony 1, 2, 3 in lane 1 - 3 respectively. 1hr after induction with IPTG for colony 1 - 3 in lanes 4 -6 respectively. Over expression of MalE is marked out in red dashed box in lanes 4 - 6. Peptide markers are in lane 0 in both panel A and B, with molecular weights marked out in kDa at respective border. Overexpression for the three different scaled up cell cultures, after lysis, can be seen in panel A in lane 7 - 9, non-induced, and in lanes 10 - 12, after 3 hr induction. MalE over expression marked out in red dashed box in lanes 10 - 12. Ni-NTA-chromatography is shown in panel B. Lane 1 and 2 shows the flow through, lane 3 and 4 shows the first 2 washing steps with buffer without Imidazole. Lane 5 - 7 shows washing steps with buffer with 25 mM Imidazole added to prevent any unspecific binding. Elution of MalE with buffer, with 250 mM Imidazole added, can be seen in lanes 8 - 11. Eluted MalE is marked out with red dashed box in lanes 8 - 11.

After successful Ni-NTA elution the protein were subjected to gelfiltration. The FPLC chromatograms show a single sharp peak of MalE (Figure 2). MalE in 20 mM phosphate 100 mM NaCl buffer (pH 7.4) were stable for seven days at room temperature at concentrations ranging between 0.1 - 2.5 mM. Protein solutions up to 3mM were stable at same conditions for two - three days.

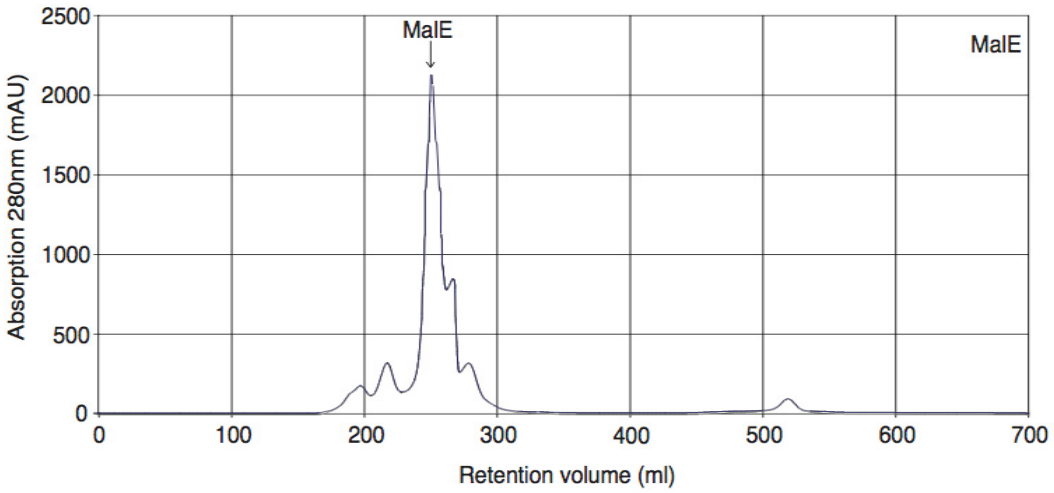

Figure 2.1.2.2. FPLC chromatogram MalE
Gelfiltration chromatography of MalE subjected on a Superdex 75, 500ml gelfiltration column on a FPLC. MalE comes with a retention volume of 254 ml. The small shoulder to the right of the MalE is most likely due to overload of the column.

2.2 Characterization of the soluble protein MalF-P2 in solution

Analytical methods were used to provide initial data regarding fold, size and any oligomeric species of the proteins in solution. Experiments demonstrate that MalF-P2 is a well-folded 21kDa monomeric protein even at high concentrations in solution.

2.2.1 Mass spectrometry (MS)

Protein samples of MalF-P2 and MalF-P2-His_6 were checked after expression and purification with mass spectrometry (Figure 2.2.1.1) using MS/MS. His_6-tag cleavage was also verified by detection of the smaller molecular weight and presence of the His_6-tag alone. MalF-P2 with and without the His_6-tag matches the calculated molecular weight of the polypeptide (Table 2.2.1.1). The correct primary sequence was also determined by in-gel trypsin digestion[96].

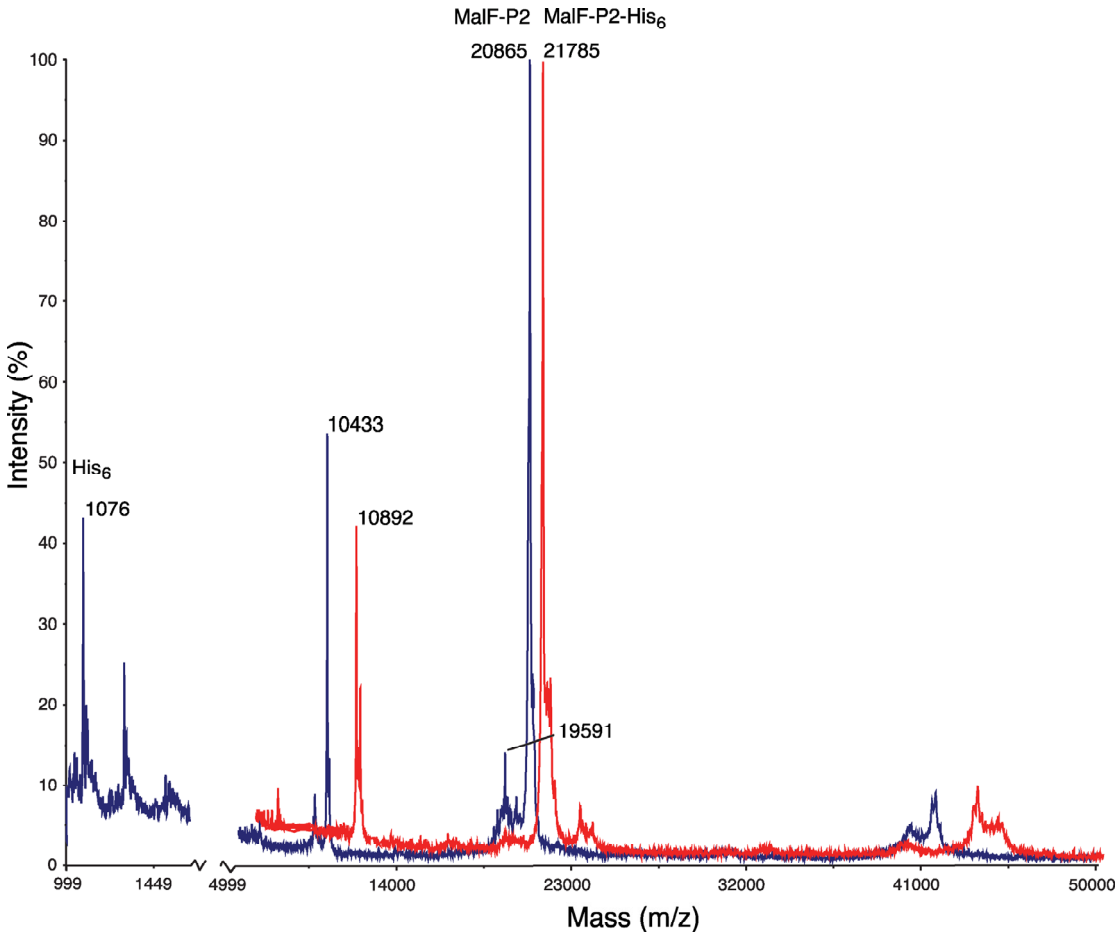

Figure 2.2.1.1. Mass spectrometry chromatograms of MalF-P2-His_6 (red) and MalF-P2 (blue).
Sharp single peaks can be seen for MalF-P2-His_6 (red) and MalF-P2 (blue) at 21785 and 20865 Da respectively. Satellite peaks of double charged species are observed for both species at half the mass 10892 and 10433 Da, respectively. The thrombine cleavage of His6 can be confirmed with a peptide peak at 1076 Da (blue spectra).

Two samples of MalF-P2 were further analysed by in-gel trypsin digestion. The MALDI-TOF/TOF data of the in-gel trypsin digested protein were analysed with MASCOT[97]. The two digested MalF-P2 samples received a MASCOT score of 137 and 286 respectively, with 8 and 12 peptides detected, where 2 and 5 were verified with MS/MS. Sequence coverage of 63 and 71 %, respectively, verified the expected sequence.

Table 2.2.1.1. Molecular mass predictions and MS determined molecular mass MalF-P2.
Molecular mass predictions and experimentally determined values by MS/MS measurements for MalF-P2 with and without His_6-tag.

Polypeptide	Calculated Mass (Da)	Measured mass (Da)
MalF-P2-His_6	21503.7	21785.1
MalF-P2	20572.7	20865.4

In-gel trypsin digested protein was also used to determine the oxidation state of Met91, Met197 and Met198. This was not successful due to the ion-spray sample loader. All three methionines were randomly oxidized and the level of oxidation on the different methionines could not be quantified.

2.2.2 Analytical ultracentrifugation (AUC)

Analytical ultracentrifugation experiments were conducted for MalF-P2 and MalE to retrieve any oligomeric states of the proteins in solution. From the direct boundary modelling of data from the MalF-P2 and MalE samples (Figures 2.2.2.1-2), a c(s) distribution of Lamm's equation was calculated that gives the sedimentation velocity data. The continuous sedimentation distribution c(s) with maximum entropy regularization shows only one sharp peak for the MalF-P2 and MalE samples (Figure 2.2.2.1 panel B and C for MalF-P2 and Figure 2.2.2.2 panel B for MalE, respectively). After transformation to a molar mass distribution and an area determination of peaks in the spectra, one finds a single peak (98% and 97% of total area for the experiment of MalF-P2 and MalE respectively) at 21 kDa for MalF-P2 (Figure 2.2.2.1 panel B and C) and 42 kDa for MalE (Figure 2.2.2.2 panel B). Identical results were obtained for experiments with different concentrations (100, 200 and 500 mM respectively) of MalF-P2 and MalE (data not shown). AUC experiments were performed together with Nikolai Wenta (FMP, Berlin).

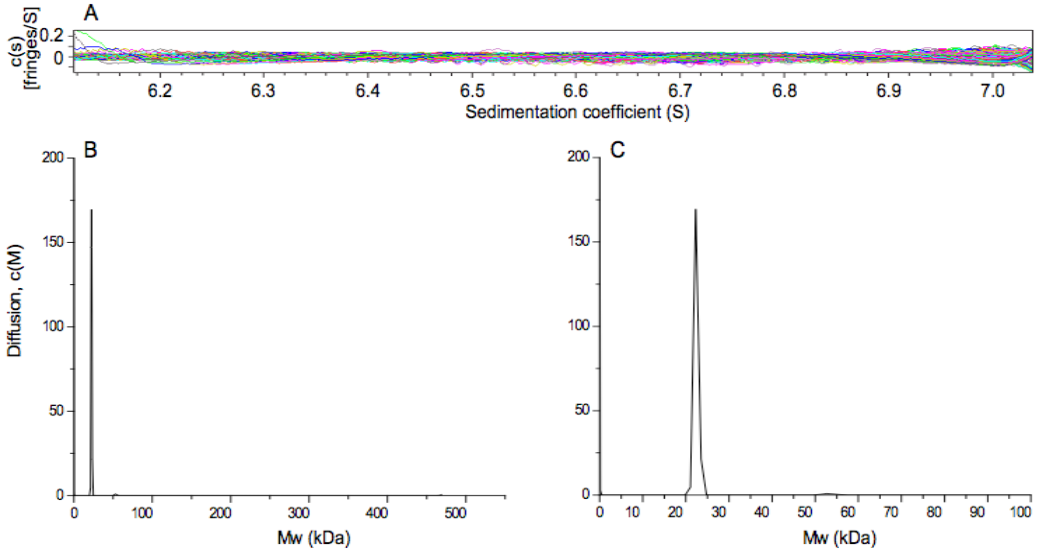

Figure 2.2.2.1. Analytical ultracentrifugation data for MalF-P2.
Panel A shows the sedimentation velocity data. The continuous sedimentation distribution transformed into molar mass distribution is shown in panel B and C at two different molecular weight regions, 0 – 550 kDa in A and 0 – 100 kDa in B. Panel B also shows the raw sedimentation data of the c(s) distribution. Only one single sharp peak can be detected in both molecular weight regions, indicating that MalF-P2 is monomeric.

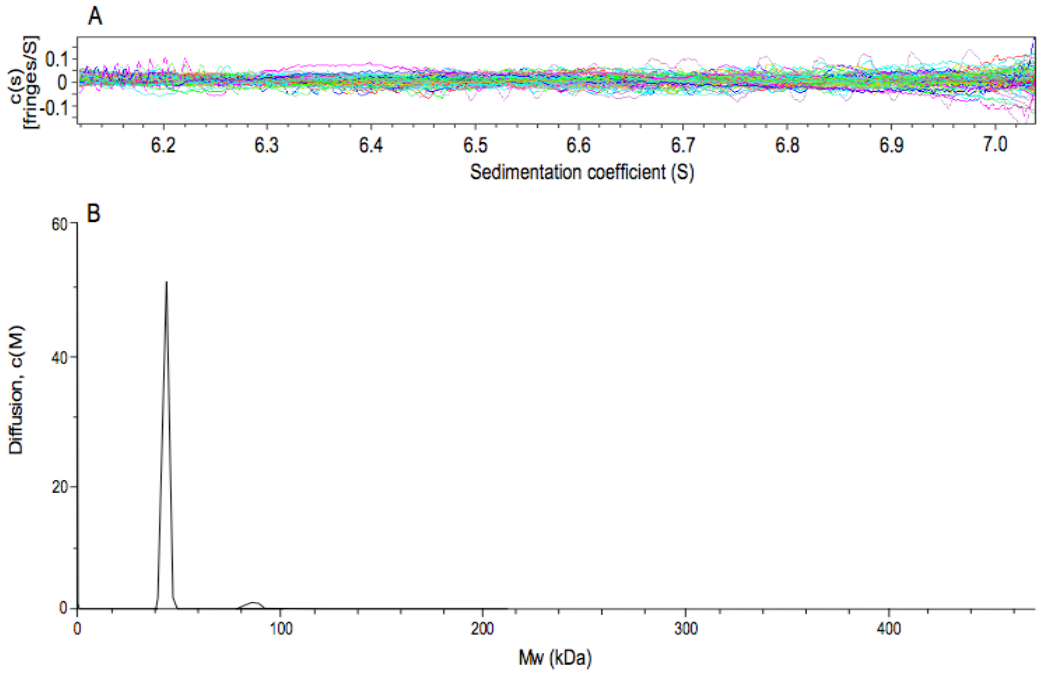

Figure 2.2.2.2. Analytical ultracentrifugation data for MalE.
Panel A shows the sedimentation velocity data. The continuous sedimentation distribution transformed into molar mass distribution is shown in panel B in a region between 0 – 450 kDa. Panel B also shows the raw sedimentation data of the c(s) distribution. Only one single sharp peak can be detected in the molecular weight region, indicating that MalE is monomeric.

Sedimentation runs for the MalF-P2 and MalE complex were not successful and gave no clear data. NMR interaction data and ITC, presented in section 2.5, can in this case provide sufficient data to determine the interaction between the proteins to a molar ratio of 1:1.

2.2.3 Circular dichroism (CD-) spectroscopy

The secondary structure of MalF-P2 in solution was examined by CD (circular dichroism) spectroscopy. Computer analysis of the CD-spectrum (Figure 2.2.3.1 panel A) revealed an approximate content of 4.3% α-helices, 55% β-strands and 40% of random coil structure (computer analysis performed by Mathias Grote). No significant changes could be detected after removal of the His_6-tag, data not shown. After addition of equimolar amount of MalE the computer analysis revealed a content of 12.7% α-helices, 54.1% β-strands and 33.3% of random coil structure after the MalE spectrum had been subtracted.

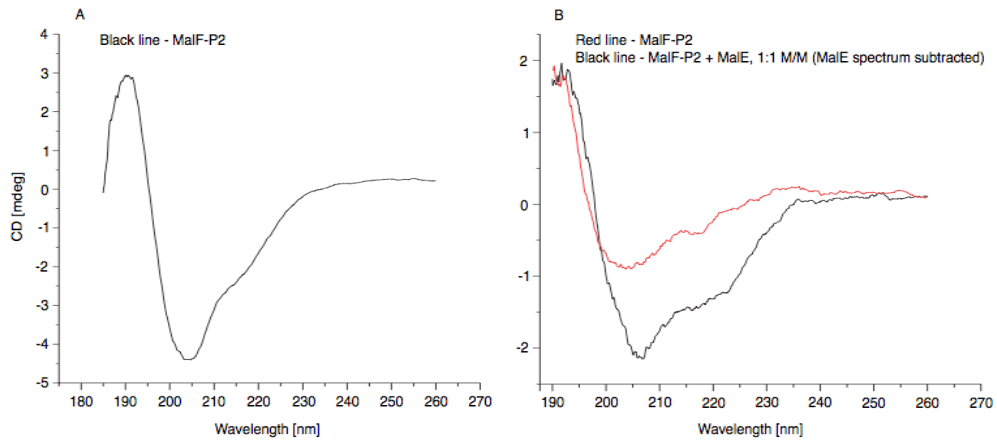

Figure 2.2.3.1. CD spectrum of MalF-P2 and MalF-P2/MalE.
CD spectrum for secondary structure of MalF-P2 and changes to it due to protein-protein interactions. Panel A shows the CD data for MalF-P2, the computer analysis revealed a content of 4.3% α-helices, 55% β-strands and 40% of random coil structure. The CD spectra of MalF-P2 were re-recorded with equimolar amounts of MalE. The reference MalE CD spectra were subtracted and revealed a 8% change in α-helical secondary structure upon titration, β-strand secondary structure remained the same. Spectra were recorded with 100 uM of MalF-P2, panel A, and equimolar amounts of MalE upon titration, panel B, with a buffer containing 10 mM Na-phosphate and 100 mM NaF at pH 7.4.

The subtle change in secondary structure after addition of MalE of equimolar amounts give indication of a protein-protein interaction. This interaction was confirmed by ITC and NMR data shown in section 2.5. Even though the CD data is not a very accurate measure of determining protein-protein interactions it gave the motivation to continue.

2.3 Assignments of the soluble protein MalF-P2 by NMR

Resonance assignments of the expressed soluble protein MalF-P2 were determined in solution. Backbone assignments were used to characterize the binding interface of MalF-P2 to MalE and also to classify secondary structure elements of the polypeptide. Sidechain assignments were used for structural calculations and to assign tryptophan sidechains in the ^{1}H-^{15}N HSQC spectrum. Backbone assignments could be determined to 92% and sidechain assignments to 56%. Assignments were deposited into the BioMagResBank under accession number BMRB-15911.

2.3.1 Backbone assignments MalF-P2

Small proteins (<10 kDa) can be easily assigned by using traditional experiments like NOESY, COSY and TOCSY. This approach suffers heavily from the multiple assignment possibilities in the recorded spectra. A much better approach is to use resonance assignment methods based on triple resonance experiments. Introduction of a third dimension makes overlap of resonances much less frequent. For larger proteins (>20 kDa) relaxation starts to be a problem for these longer experiments. The T_2-relaxation is so short that most of the magnetization is lost during the pulse sequence. By adding deuterons to the carbon chains, the T_2-relaxation can be greatly prolonged.

With this approach the standard triple resonance experiments can be used, but in a slightly modified fashion. Another troublesome property of the short T_2-relaxation is that the linewidth depends directly on it. The linewidths become very large, decreasing the resolution and signal-to-noise for the spectra. By use of modern Trosy techniques[98] where only the most favourable T_2-relaxation occurs, the line broadening can be dealt with. The MalF-P2 carbon chain required deuteration for more elaborate pulse programs to work[99]. Multiple pairs of these correlation experiments[100] were used for removing the ambiguity or loss of information from one set of experiments. All spectra were assigned using the software CCPN[101].

Sequence specific backbone $^{1}H^{N}$, ^{15}N and ^{13}C chemical shift assignments of MalF-P2 were obtained using HNCOCACB, HNCACB and HNCOCA, HNCA experiment pairs[100]. The ^{1}H, ^{15}N, ^{13}C resonance backbone assignment for MalF-P2 are complete (90% N^{H}, 94% Cα Cβ) with exception of two segments, one being the first eight residues in the N-terminus and the other containing residues 167 – 177 in the C-terminal region. In the X-ray structure of the full maltose transporter this region makes an amphipatic helix to the lipid membrane[45]. This helix can most probably not fold up properly or is in an equilibrium switching between different structural states, making the assignments inconsistent in this section.

Characteristic downfield Cβ chemical shifts of threonine and serine residues and upfield chemical shifts of the alanine Cβ's served as starting points for the assignment (Figure 2.3.1.1). By walking along the protein backbone the resonance assignments could be achieved by using the HNCOCACB, HNCACB experiment pair, the assignments were verified by using the HNCOCA, HNCA experiment.

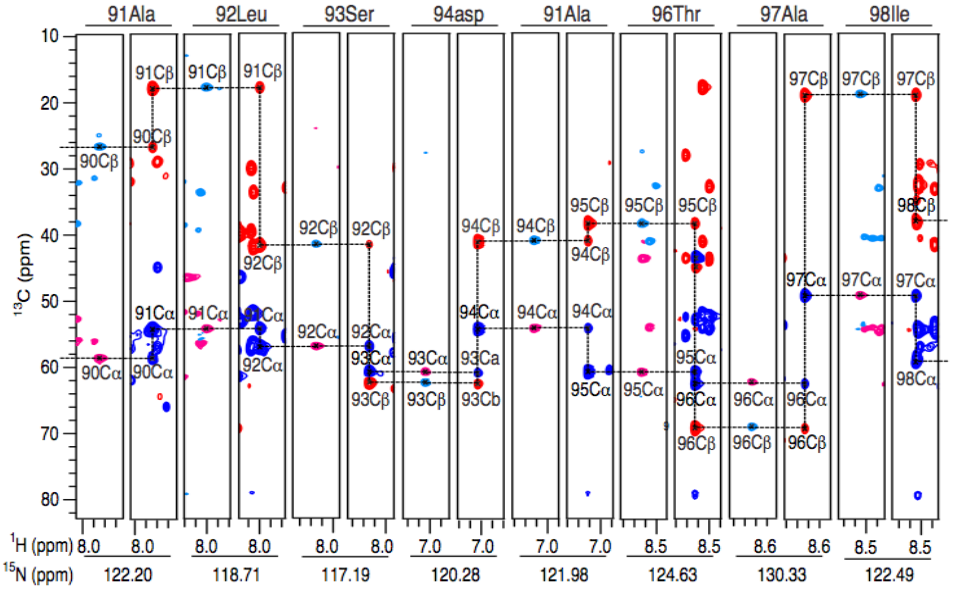

Figure 2.3.1.1. Strip plots of backbone assignments of MalF-P2

Strip plots of the HNCOCACB (left column of each residue) and HNCACB (right column of each residue) spectra. Sequential connectivities for the protein backbone were obtained by connecting the Cα (red peaks, negative intensity) and Cβ (blue peaks, positive intensity) signals of the i-1th residue in the HNCOCACB with the Cα (blue peaks, negative intensity) and Cβ (red peaks, positive intensity) of the ith and i-1th residue in the HNCACB experiments.

Figure 2.3.1.2 shows the $^{1}H^{N}$, ^{15}N chemical shifts for the backbone resonances and tryptophan indoles in a ^{1}H, ^{15}N-HSQC-Trosy spectrum of MalF-P2.

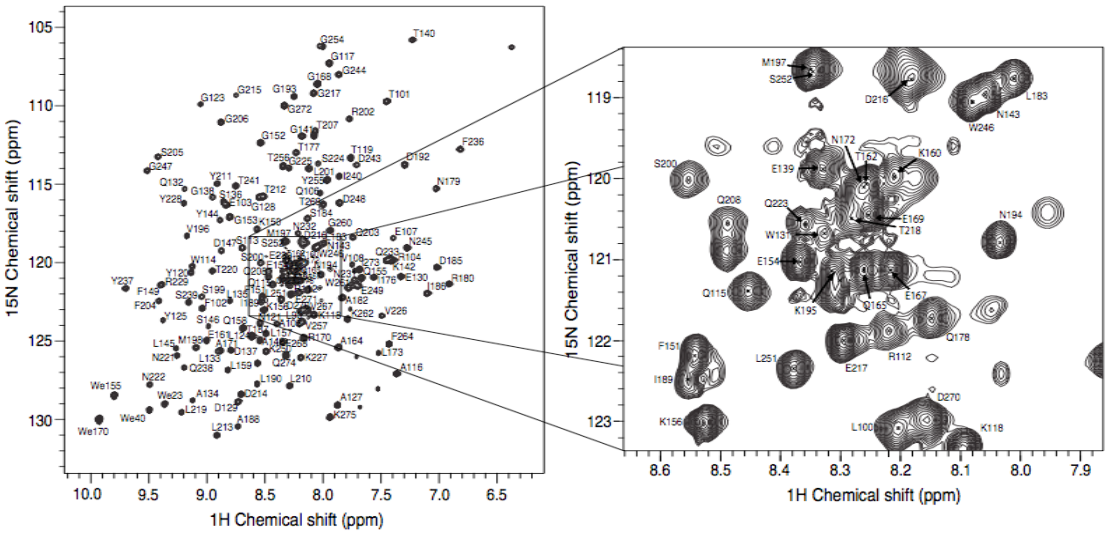

Figure 2.3.1.2. ^{1}H-^{15}N HSQC-Trosy spectrum of MalF-P2.
^{1}H-^{15}N HSQC-Trosy spectrum of MalF-P2 at a temperature of 300K on a Bruker DRX 600 spectrometer. Assignments for the $^{1}H^{N}$, ^{15}N chemical shifts for the backbone resonances and tryptophan indoles are marked out in the spectra. The overlapped mid region of the spectra has been zoomed in to highlight assigned residues in that area.

Backbone chemical shift assignments for MalF-P2 are deposited in the BioMagResBank, http://www.bmrb.wisc.edu/, under the accession number BMRB-15911.

2.3.2 Sidechain assignments of MalF-P2

Following the backbone resonance assignments, the assignment of the ^{1}H and ^{13}C nuclei of the amino acid side-chains was carried out. Aliphatic side-chain assignments for MalF-P2 were achieved by standard 3D experiments for non-deuterated proteins[100]. ^{13}C chemical shifts were corrected for the deuterium effect according to Venters *et al.*[99]. Hα chemical shifts of assigned amino acids could be extracted from a HACACONH experiment to correlate the Hα chemical shifts to the deuterium corrected Cα chemical shifts in the 3D HCCH-COSY and HCCH-TOCSY. Using the 3D HCCH-COSY and 3D HCCH-TOCSY (Figure 2.3.2.1) 56% of the side-chain assignments could be determined. The lower number of side-chain assignments was due to large ambiguities of flexible side-chains that overlap in the spectra. Some ^{13}C chemical shift differences at the γ position and out from the side-chain could also be seen between different samples. This emphasis the dynamics of the side-chains of MalF-P2 in solution, where it can accommodate different rotamers, hence distorting the side-chain chemical shifts.

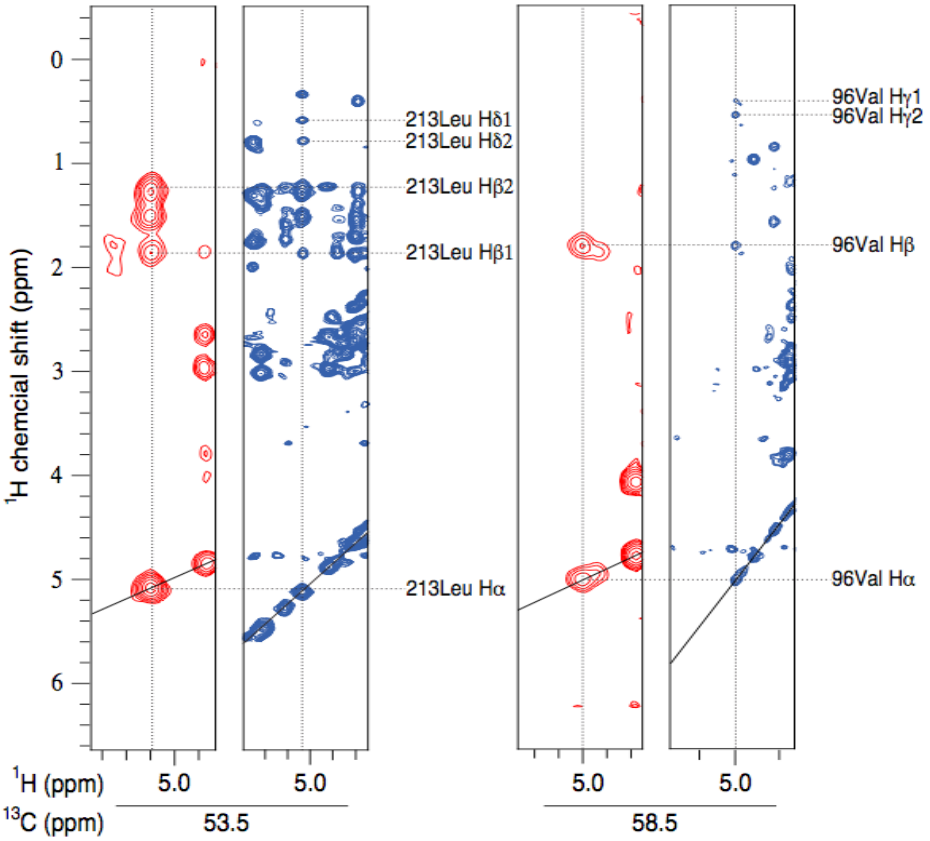

Figure 2.3.2.1. HCCH-TOCSY and HCCH-COSY strip plots sidechain assignments of MalF-P2. HCCH-COSY (red resonances) and HCCH-TOCSY (blue resonances) experiments for the side-chains for the amino acids Leu 213 (left two strips) and Val 96 (right two strips), referring to Hα of the amino acid shown in the x and y-axis and the corresponding Cα chemical shift in the z-axis. Correlating peaks for ^{1}H chemical shifts per residue in the COSY and TOCSY experiments are marked with dashed lines.

A 3D ^{1}H-NOESY-^{1}H,^{15}N-HSQC experiment was used to assign the indole $^{1}H^{N}$, ^{15}N chemical shifts of the tryptophans (Figure 2.3.2.2) by correlating NOE peaks from the indole ^{1}H, ^{15}N chemical shifts to the backbone ^{1}H, ^{15}N chemical shifts of the tryptophans.

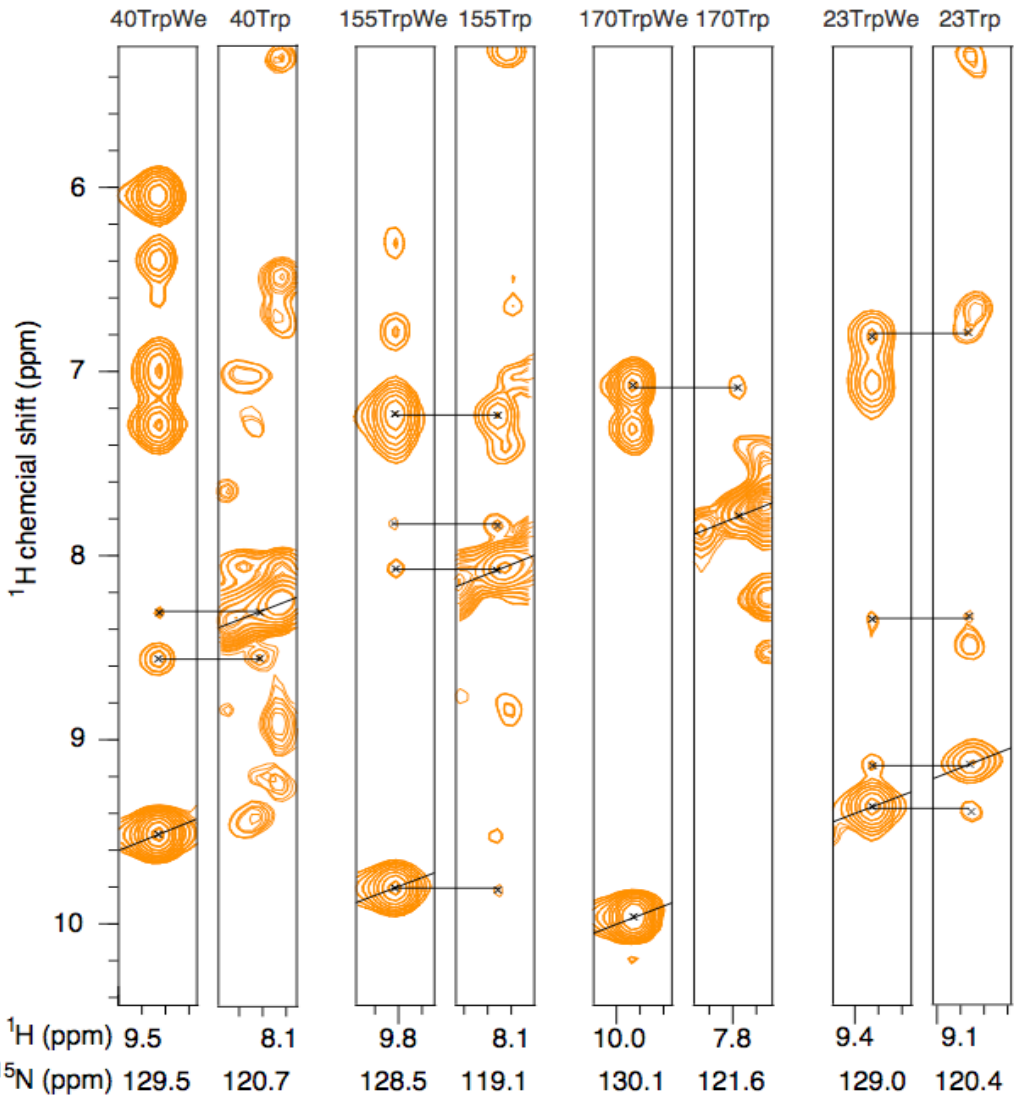

Figure 2.3.2.2. ^{1}H-NOESY-^{1}H,^{15}N-HSQC strip plots for indole group assignments of MalF-P2.
3D ^{1}H-NOESY-^{1}H,^{15}N-HSQC experiment for assignment of the ^{1}H, ^{15}N chemical shifts of the tryptophan indole groups. Wε 40, Wε 155, Wε 170 and Wε 23, left to right. ^{1}H, ^{15}N chemical shifts of the four indole groups ($^{1}H^{N}$ x-axis, $^{1}H^{N}$ y-axis, ^{15}N z-axis), shown side by side left to right. Correlating ^{1}H NOE peaks between the indole and backbone chemical shifts are connected by lines, confirming the assignment.

2.4 Structural characterization of MalF-P2 in solution by NMR

NMR yields the secondary structure of a protein at an early stage in the structural investigation. Through analysis of chemical shifts different methods can correlate secondary structure elements of the studied protein[102,103,104].

Cα, Cβ chemical shifts that deviates from random coil chemical shifts can be used to identify α-helical and β-strand regions of the protein. Experimentally determined Cα and Cβ shifts are subtracted from tabulated random coil chemical shifts for the corresponding amino acids, α-helical regions are characterized by increased Cα and decreased Cβ chemical shifts, and β-strands with decreased Cα and increased Cβ

chemical shifts. An α-helix needs to have four subsequent α-helical chemical shifts in order to be properly identified. β-strands need three subsequent residues to be considered a β-strand.

For MalF-P2 two α-helices (between residues 100–109 and 172-179) and nine β-strands (residues 120–127, 131–139, 141–148, 192-194, 201-204, 210-212, 216-219, 237-239 and 247-249) can be identified from the chemical shifts analysis (Figure 2.4.1).

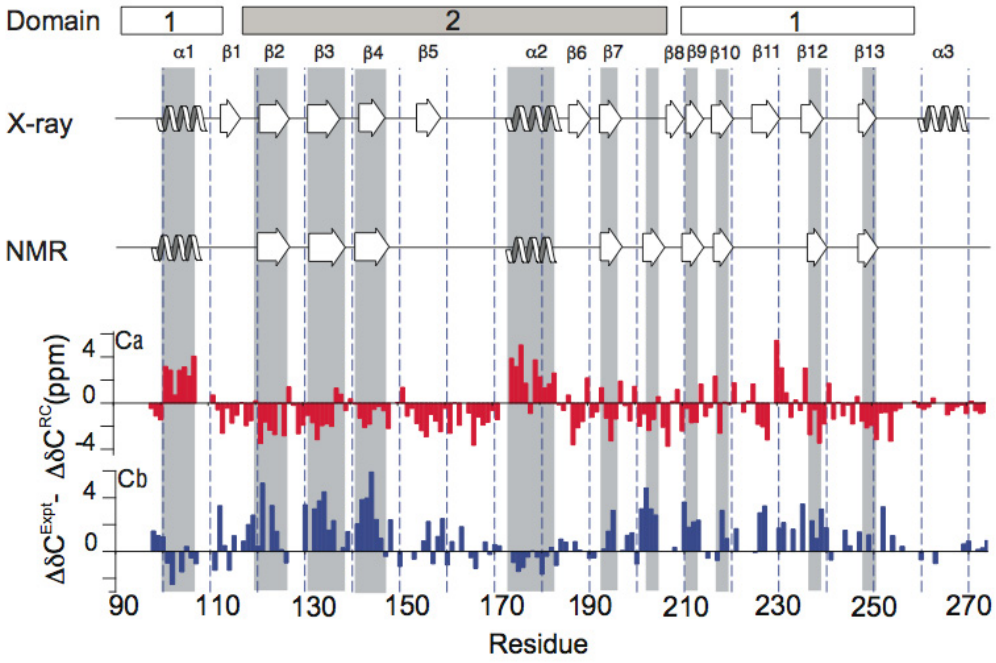

Figure 2.4.1. Comparison of secondary structure elements. NMR and X-ray, for MalF-P2.
Comparison of adopted secondary structure elements of MalF-P2 in the $MalFGK_2$-E X-ray structure[45] and in the separately expressed MalF-P2 loop as predicted by NMR chemical shift analysis. The indicated secondary structure is based on the Cα, Cβ NMR chemical shifts data employing the program TALOS[104]. The bottom of the Figure represents differences between random coil (RC) and experimental (expt) chemical shifts for Cα and Cβ as a function of the primary sequence in MalF-P2. Secondary structure elements as predicted by TALOS are highlighted in grey. MalF-P2 folds into two individual domains (domain 1: residues 91-113 and 209-259; domain 2: residues 117-207), which are connected via the β-strands β1 and β2 from the crystal structure.

Dihedral angles Φ and Ψ of the peptide bond can be predicted based on Cα, Cβ, N, N^H chemical shifts of all the assigned non-proline residues. MalF-P2 chemical shift data were predicted by using the program TALOS[104]. Several values were predicted for each residue by TALOS and the predictions were checked visually based on the distribution of prediction in the Ramachandran plot. Only the angles, which classified as "good" by the program were used in further calculations. All the torsion angles from TALOS were used in the calculations with an error margin of $\pm 30^0$. Torsion

angles predicted by TALOS are presented in Figure 2.4.1 together with the secondary structure picked by the *dss* algorithm in PyMol[55].

Both chemical shift analysis and TALOS secondary structure prediction are in good agreement with the crystal structure[45]. α-helix 3 does not show any α-helical propensity in solution. In the crystal structure this α-helix adopts an amphipathic helix anchored into the lipid membrane. Backbone resonances could also not be fully assigned in this region, most likely due to the dynamic character of this region in solution, giving gaps of missing chemical shifts. β-strand 1 and 11 are missing in the chemical shift analysis and TALOS prediction. These two regions are assigned as β-strand secondary structure from the *dss* algorithm in PyMol[55] even though they only consist of two consecutive amino acids.

2.5 Interactions of MalF-P2 and MalE in solution

The interaction of MalF-P2 to MalE was characterized by biochemical and biophysical experiments. Chemical cross-linking experiments were conducted to show consistency of the soluble MalF-P2 and MalE interaction in solution to the interaction of MalE and the assembled $MalFGK_2$ transporter. The strength of the interaction was measured by ITC experiments in presence and absence of substrate for MalE. Solution NMR titration experiments characterized the binding interaction and mode between the proteins. These were then later mapped onto the interactions as seen in the crystal structure. Through these experiments a detailed picture of the interactions between MalF-P2 and MalE could be obtained.

2.5.1 Chemical cross-linking with sulfonate cross-linkers

Earlier work performed by Daus *et al.*[96], involves studies of the full catalytic cycle of the full maltose transporter $MalFGK_2$-E. Interaction sites between the different transmembrane domains were mapped out at different times of the transport cycle[96,105,106]. The function of the P2 loop of the MalF transmembrane domain could however not be fully understood. A more detailed study of the interaction between MalF-P2 and MalE was required. MalF-P2 was therefore expressed separately (see Materials and methods section 4.5 and 4.8) to confirm directly the interaction of purified MalF-P2 and MalE. Cross-linking experiments were performed for this purpose (experiments performed by Mathias Grote, HU Berlin). The mutation points on MalE and MalF-P2 were taken from the earlier work of Schneider *et al.*[105,106]. A sample with a variant of MalF-P2 containing a T177C mutation was incubated with MalE(T31C) at equimolar (1:1) molar ratio in the presence of CuPhe, and subsequently analyzed for disulfide-bond formation by SDS-PAGE. Besides monomers and homodimers of both proteins, an additional band, comprising MalE and MalF-P2 could be observed (Figure 2.5.1.1). This was verified by immunoblotting (data not shown). The product was found independent of the presence or absence of maltose, thereby confirming the NMR results and the ITC interaction studies, see sections 2.5.2-3. In a control experiment, no cross-link product

was found with MalF-P2(T177C) and MalE(K179C) (data not shown), which is in accordance with the crystal structure of the maltose transporter by Oldham *et al.*[45] This result underscores the specificity of the MalF-P2 - MalE interaction.

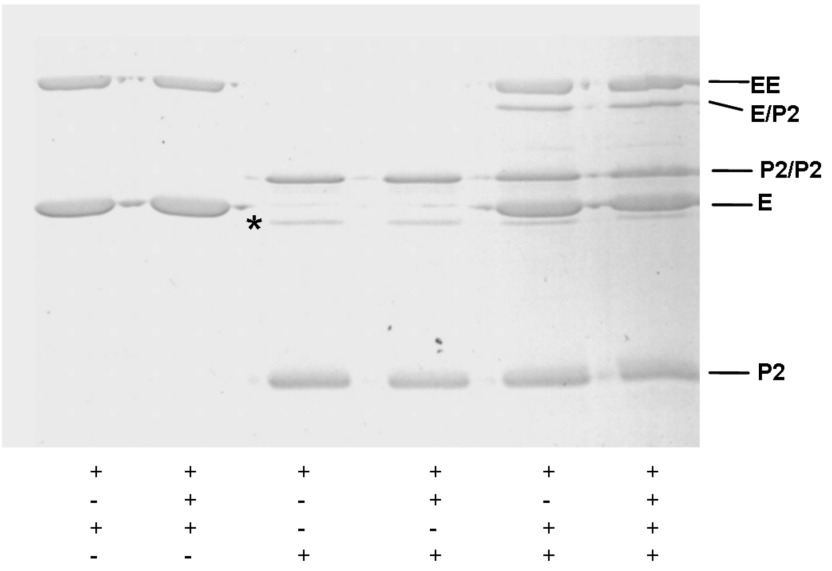

Figure 2.5.1.1. Site-specific CuPhe-induced cross-linking of MalF-P2.
Site-specific CuPhe-induced cross-linking of the purified MalF P2-loop carrying a cysteine at position 177 with MalE(T31C). MalF-P2 (2.5 μM) was incubated with MalE(T31C) (2.5 μM) and CuPhe in the absence or presence of maltose. The band marked with an asterisk is an impurity and is co-purified only with the monocystein variant of MalF-P2. Figure from Jacso *et al.* 2009[107] (experiments performed by Mathias Grote, HU Berlin).

This also, together with earlier work from Schneider *et al.*, confirms the interaction seen between the expressed variant of MalF-P2 and MalE, and is consistent with an interaction between MalE and the assembled $MalFGK_2$ complex throughout the transport cycle[108].

2.5.2 Isothermal Titration Calorimetry (ITC)

In order to find out to what extent MalE can interact with the isolated P2 loop of MalF, isothermal calorimetric titration (ITC) experiments were carried out. Figure 2.5.2.1 shows that there is a relatively strong interaction between MalE and MalF-P2, even in the absence of other periplasmic loops of MalF and those of MalG. Assuming a 1:1 stoichiometry, fitting of the data yielded a dissociation constant K_D on the order of 10-30 μM, independent of the presence of maltose.

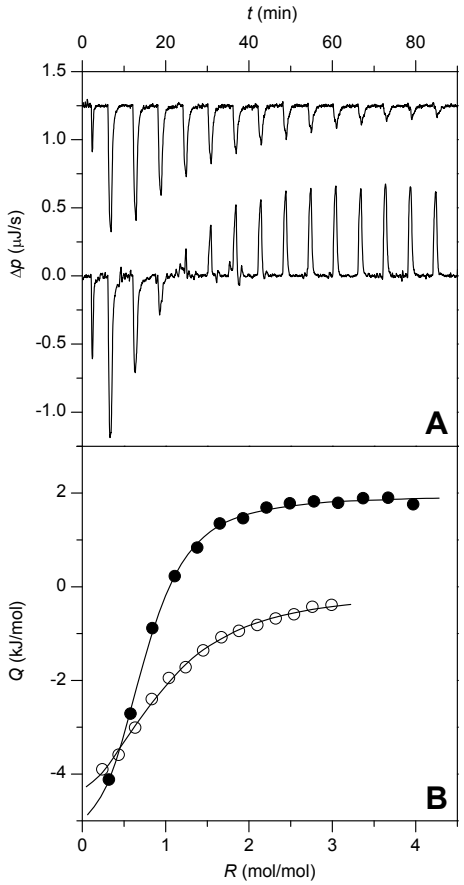

Figure 2.5.2.1. ITC measurements of MalF-P2 and MalE.
Isothermal titration calorimetry (ITC) curves obtained by titrating MalF–P2 to MalE. (A) Differential heating power, Δp, versus time, t. ITC data were recorded at 5^0C in the absence (top) and presence (bottom) of maltose. One 5-μL and 14 20-μL aliquots of a concentrated MalE solution (667 μM in the absence of maltose, 885 μM in the presence of maltose) were injected into a 50 μM MalF-P2 solution in the absence (upper trace) or presence (lower trace) of 100 mM maltose. For clarity of presentation, the ITC curve obtained without maltose (upper trace) was arbitrarily shifted along the ordinate by 1.25 μJ/s. Constant endothermic peaks at the end of the titration with maltose (lower trace) were due to dilution effects (most likely originating from slight differences in maltose concentration) and were also observed in control experiments injecting MalE into dialysis buffer containing 100 mM maltose but no P2 (data not shown). (B) Integrated and normalized heats of reaction, Q, versus molar MalE/P2 ratio, R. Fits (solid lines) to experimental data obtained in the absence (open symbols) or presence (filled symbols) of 100 mM maltose yielded the following thermodynamic parameters characterizing the MalF–P2 interactions: K_D = 22 μM, ΔG° = –24.9 kJ/mol, ΔH° = –6.2 kJ/mol, and TΔS° = 18.6 kJ/mol in the absence of maltose; K_D = 7.4 μM, ΔG° = –27.3 kJ/mol, ΔH° = –8.5 kJ/mol, and TΔS° = 18.8 kJ/mol in the presence of 100 mM maltose. Initial 5-μL injections were excluded from the fitting procedure.

Differences in the apparent affinity between MalF-P2 and MalE in absence ($K_D \approx 22$ μM) and presence ($K_D \approx 7.4$ μM) of maltose are most likely due to experimental error. The endothermic behaviour in the presence of maltose at the end of the titration is due to the heat of dilution, which would originate if the concentrations in the syringe and the ITC chamber were slightly different.

Experiments with MalF-P2 and MBP were also conducted at higher temperature (25^0C) to find optimal the ΔH for the reaction. Similar results as above were obtained. We find a less pronounced interaction at higher temperature.

Control experiments were performed to see if maltose alone interacts with MalF-P2. Experiments were performed with MalF-P2 in the measuring cell and maltose in the injection syringe. From the titration with MalF-P2 and maltose there is a clear indication that maltose alone does not interact with MalF-P2. Blank experiments with buffer injected into either MalF-P2 or MBP showed no significant and only minor endothermic behaviour.

2.5.3 Titration experiments with MalF-P2

The interaction between MalF-P2 and MalE was followed with site specific resolution using ^{1}H-^{15}N HSQC experiments (Figure 2.5.3.1-2). If MalE is added to a solution of ^{15}N isotopically enriched MalF-P2, specific resonances display decreased intensities or perturbations. Spectra in presence and absence of MalE were recorded under identical conditions (identical buffer pre-equilibrated with or without maltose). Similar chemical shift perturbations are observed in the presence of maltose. This behaviour is expected as the observed dissociation constant of the binding is in the intermediate micromolar range. In case of strong binding, two separate sets of resonances are expected reflecting the different chemical environment of bound and free MalF-P2. In the weak-binding limit, a continuous change of the ^{1}H,^{15}N chemical shifts of MalF-P2 as a function of the concentration of MalE would yield the dissociation constant of the interaction. In case of intermediate binding, line broadening for those residues interacting with MalE occurs. At a molar ratio of 10:1 of MalF-P2 versus MalE,

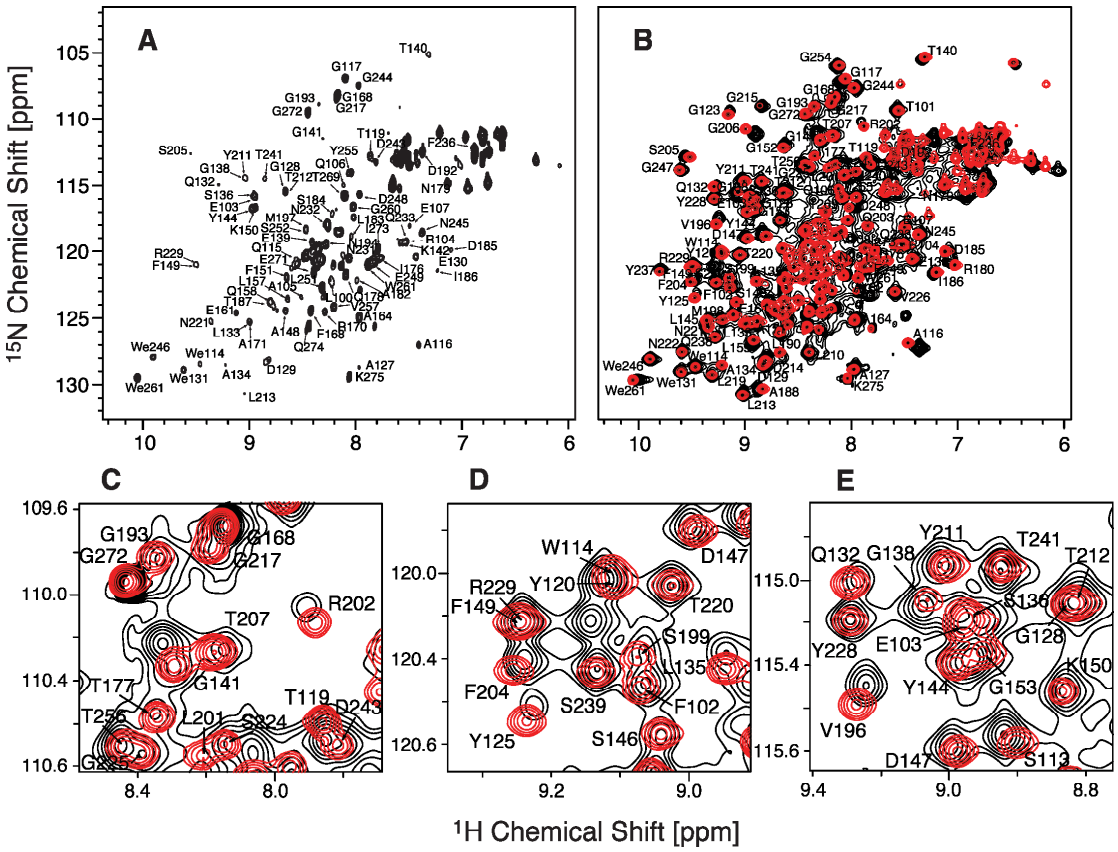

Figure 2.5.3.1. ^{1}H, ^{15}N HSQC titration experiments with MalF-P2 and MalE w/wo maltose.
A, ^{1}H, ^{15}N HSQC titration spectra of ^{15}N isotopically enriched MalF-P2 with MalE (molar ratio 1:1). B, Superposition of MalF-P2 in absence (red) and presence (black) of unlabeled MalE. In contrast to A, the spectra of MalF-P2 in presence of MalE is represented at lower contour levels. Panels C-E focuses on particular spectral regions of the spectra represented in panel B.

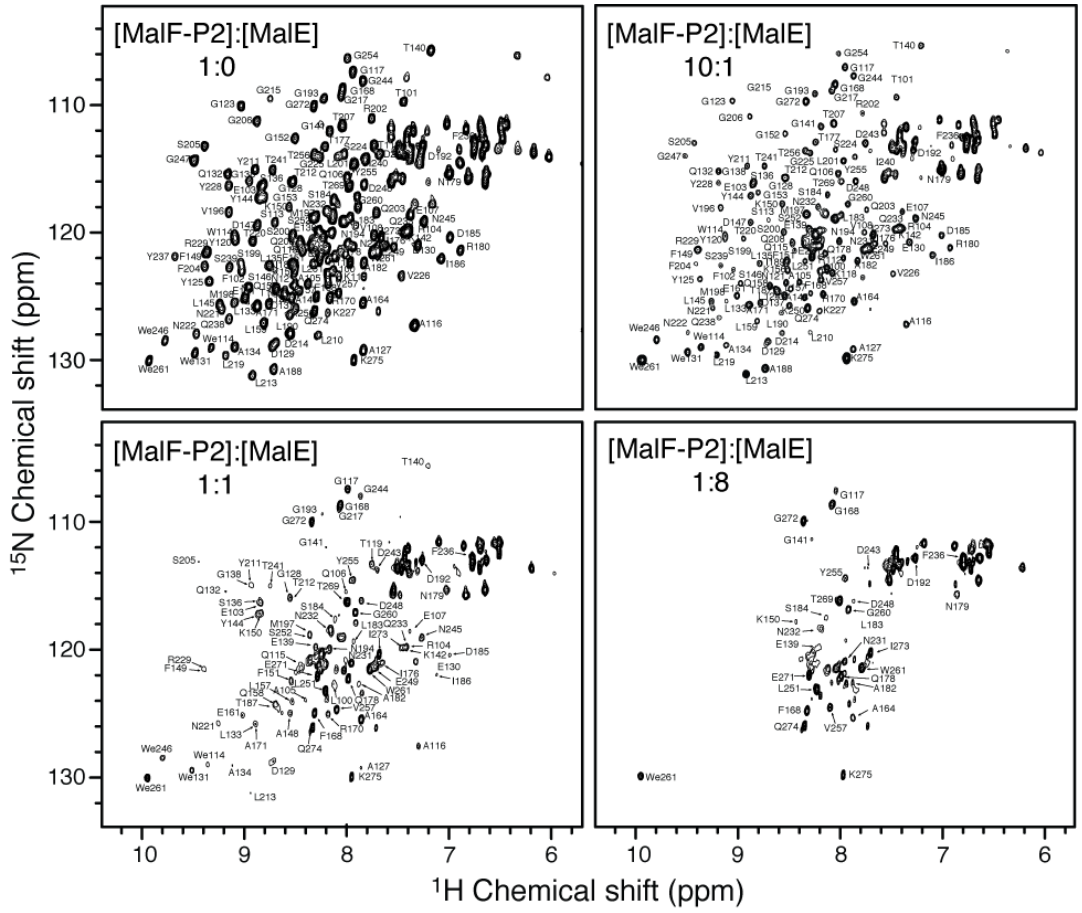

Figure 2.5.3.2. ^{1}H, ^{15}N HSQC titration experiments with MalF-P2 and MalE at different molar ratios. ^{1}H,^{15}N correlation spectrum of MalF-P2 upon titration of MalE. The ratio of the respective protein concentrations is indicated in the figure. All spectra are represented at the same contour levels.

small perturbations of the chemical shifts are detected. Already at this low concentration of MalE, we find slightly reduced intensities of the MalF-P2 resonances. At a 1:1 molar ratio, the intensity of a large number of resonances is drastically reduced as a consequence of the interaction with MalE. Residues belonging to the C terminal part and to the external loops of MalF-P2 remain largely unaffected from the binding process. Strong interactions between MalF-P2 and MalE are also observed in the crystal structure of $MalFGK_2$-E maltose transporter[45]. Figure 2.5.3.3 A, summarizes the structural contacts observed in the crystal structure and the ones found in the NMR experiment. In the analysis of the X-ray contacts, all residues are taken into account for which a heavy atom (N, C, O) of MalF-P2 is within a distance of 5 Å to a heavy atom (N, C, O) of MalE. Residues, which were selected as NMR interaction sites show a chemical shift change larger than 0.03 ppm, equation [2.5.3.1], which is equivalent to a ^{1}H shift change of 22.5 Hz at 750 MHz.

$$\delta\Delta_{tot} = \sqrt{\delta\Delta\left(^{1}H\right) + \delta\Delta\left(^{15}N\right)/5} \qquad [2.5.3.1]$$

Residues, involved in binding are located in the secondary structure elements α1/β1 (residues 100-110) and α2/β6 (residues 175-191), as well as in both linker region β1/β8, which connect domains 1 and 2 including residues 112-116 and 200-208, respectively. Figure 2.5.3.3 panel B, depicts the chemical shift perturbations on a ribbon and surface representation of MalF-P2 and MalE, respectively. For MalF-P2 residues that are colour coded in orange/red, have significant chemical shift perturbations, indicating a direct involvement of those residues in binding to MalE. Residues colour coded in blue, have chemical shifts unaffected upon addition of MalE.

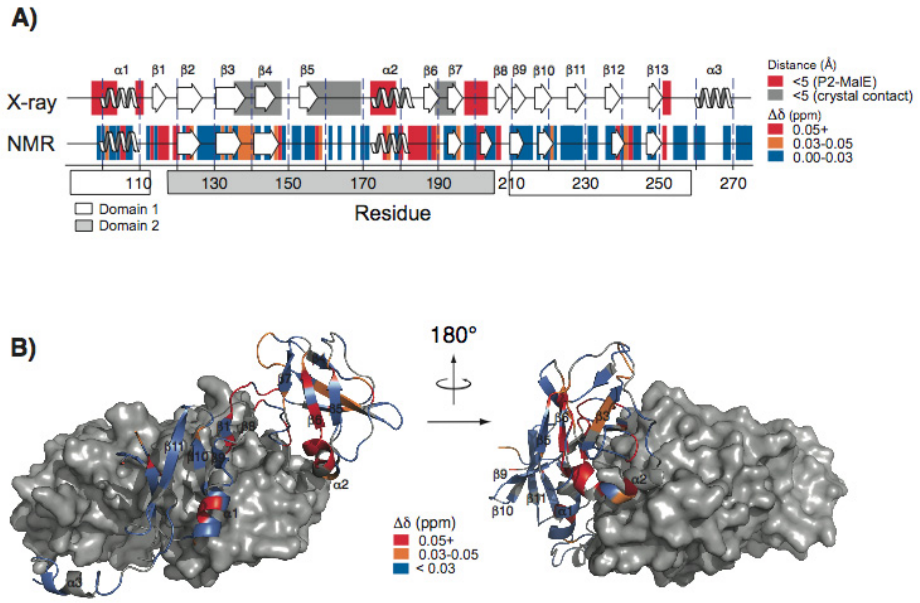

Figure 2.5.3.3. Interaction comparison between MalF-P2 and MalE from X-ray and NMR data.
Interactions between MalF-P2 and MalE, found in the crystal structure by Oldham *et al.*[45] and chemical shift perturbations. In the analysis of the X-ray contacts, all residues are taken into account in which a heavy atom of MalF-P2 is found within 5 Å to a heavy atom of MalE (marked as red boxes) or MalK (marked as grey boxes). The analysis was carried out using the program CCP4i 1.3.20[109]. Interaction sites between MalF-P2 and MalE are color coded according to the size of the perturbation. B, structural representation of the residues of MalF-P2 that are experiencing chemical shift changes, depicted as above stated, in the NMR experiment. MalE is depicted with a surface representation (grey).

2.6 Interactions of MalF-P2 and MalE as seen in the crystal structure of MalFGK$_2$-E

Structural interpretations from the crystal structure of the maltose transporter (pdb code: 2R6G) were done with CCP4i. 1.3.20 (Computational Project, Number 4. 1994[109]) and PyMol[55]. The P2-loop of MalF in the MalFGK$_2$-E complex, has two domains with the amino acid sequence folding upon itself, in a U-shape, with a short "hinge" between the two domains. Boundaries of the two domains can be set at:

domain 1: residues 91-113 and 209-259; domain 2: residues 117-207. The two regions are depicted in the left and right panels of Figure 2.6.1.

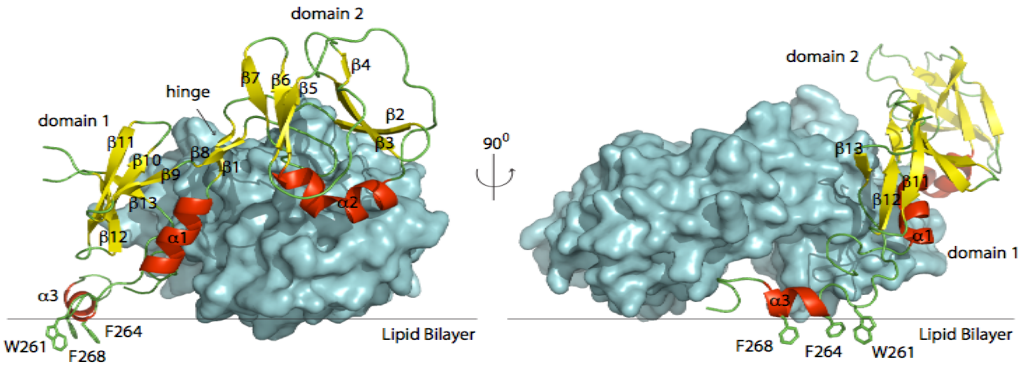

Figure 2.6.1. Complex structure of MalE and MalF-P2.
MalF-P2 and MalE complex from the $MalFGK_2$ complex analyzed by Oldham *et al.*[45] MalE is represented by a surface model (cyan) and MalF-P2 as cartoon with secondary structure elements colored in yellow (β-sheets) and red (α-helices). The figure was prepared with PyMol[55] from the $MalFGK_2$ crystal structure (pdb code 2r6g).

Secondary structure elements from the crystal structure were extracted by the *dss* algorithm of PyMol[55] (Figure 2.6.2). Secondary structure predictions by NMR for MalF-P2 in solution match the *dss* search with the exception of the hinge region, β-sheet 1 (residues 114-115) and β-sheet 8 (residues 208-209), and α-helix 3. The observed secondary structure of MalF-P2 in solution thus match the secondary structure observed in the crystal structure. β-sheet 1 and β-sheet 8 was not considered to have β-sheet propensity from the NMR due to a threshold of a minimum of three consecutive residues within the same secondary structure element. α-helix 3 can be seen "sitting" on the lipid membrane in the crystal structure with the aromatic residues W261, F264 and F268 having their aromatic regions embedded into the lipid membrane. This amphipatic α-helix is presumed not being able to fold up properly without the lipid membrane. The short hinge region and the amphipatic α-helix 3 are depicted in the left and right panels in Figure 2.6.1.

The NMR interaction of MalF-P2 in solution with and without MalE present was followed by a detailed analysis of the interaction as observed in the crystal structure to determine any inconsistencies. An atom-atom distance search were performed between MalF-P2 and MalE for distances between 0.0-5.0 Å for the heavy atoms N, C and O. 413 distances <5.0 Å could be found originating from MalF-P2 to MalE, with multiple contacts from 24 different amino acids, table 2.6.1.1. In comparison the transmembrane domain MalG has 428 contacts <5.0 Å from 27 different amino acids to MalE. The complete association between MalE and $MalFGK_2$ could thus be considered to be shared between the two transmembrane domains of MalF and MalG.

Table 2.6.1. Atom-atom distances <5.0 Å between MalF-P2 and MalE.
Atom-atom distances <5.0 Å between residues of the two domains of the P2-loop of MalF and MalE in the crystal structure of $MalFGK_2$-E (pdb: 2R6G).

Domain 1 (13 amino acids)	Domain 2 (13 residues)
Y94, S96, T97, N98, Q99, L100, T101, R104, L110, D111, R112, S252, P253	L173, R174, V175, I176, T177, Q178, R180, M198, S199, S200, L201, R202, Q203

Furthermore intermolecular interactions were analysed within the unit cell of the crystal structure of $MalFGK_2$-E. The atom-atom search yielded a large number of intermolecular distances, <5.0 Å, between domain 2 of the P2-loop of MalF to two different MalK subunits of $MalFGK_2$-E in the neighboring unit cells. Interacting amino acids of MalF-P2 are depicted in Figure 2.6.2 as shown over the primary sequence, panel A, and graphically in panel B and C.

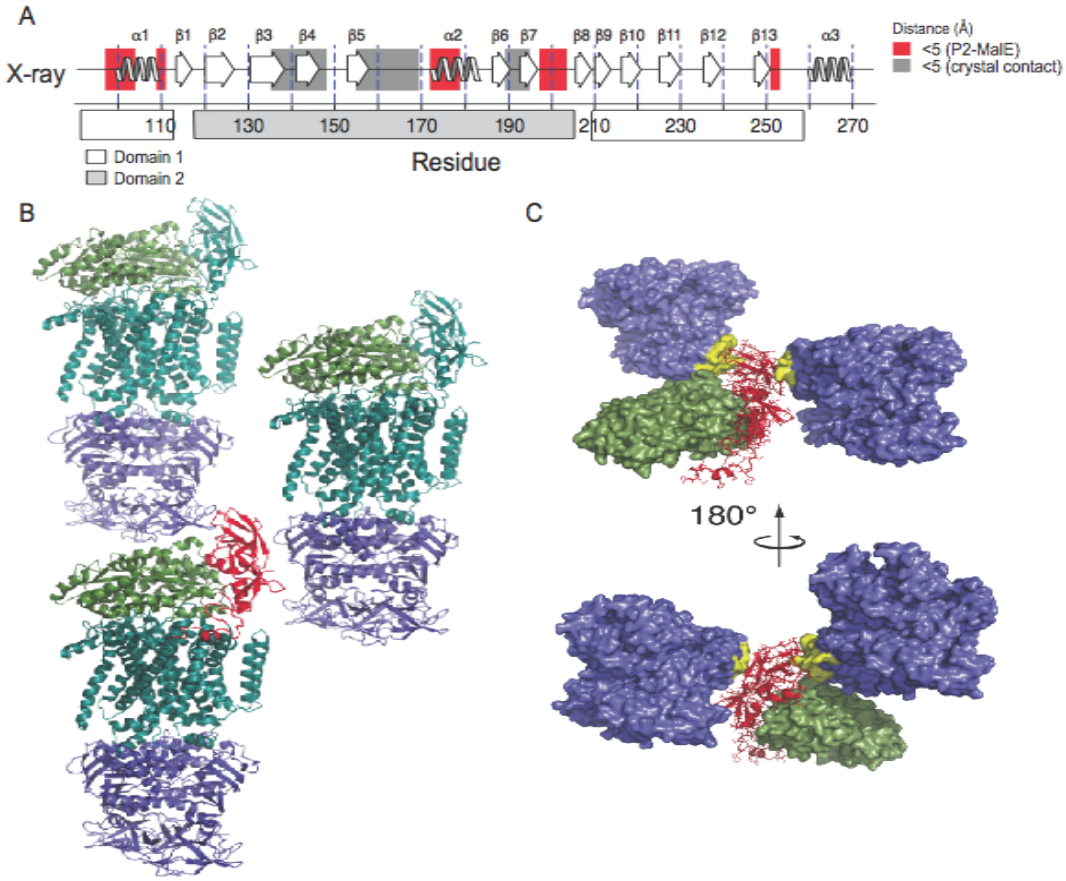

Figure 2.6.2. Crystal contacts between MalK and MalF-P2.
Crystal contacts in the X-ray structure of $MalFGK_2$-MalE (pdb: 2R6G). Intermolecular distances of <5.0 Å between MalF-P2 and MalE (red boxes) and between MalF-P2 and malK of molecules in a neighbouring unit cell (grey boxes) are highlighted. In panel B and C, MalF and MalG are drawn in dark green, MalE in light green, MalK in blue, respectively. The periplasmic loop P2 of MalF is represented in red. In particular, the second domain of MalF-P2 has many intermolecular contacts to a MalK molecule in a neighboring unit cell. Crystal contacts between MalF-P2 and MalK effects only the second domain of MalF-P2, namely the residues N121, Y125, A134, L135, S136, S138, E139, T140, G141, N143, Y144, L145, D147, E154, Q155, K156, Q158, T163, Q165, E169, R170, I189, G193, K195 (indicated in grey boxes in panel A and in yellow in the surface representation in panel C).

2.7 Structural changes of MalF-P2 upon binding of MalE

In order to probe structural differences of MalF-P2 in solution and in the crystalline state, residual dipolar coupling (RDC) measurements were performed in absence and presence of MalE and maltose[110,111]. Upon alignment, the anisotropic dipolar coupling is not averaged any longer. The residual dipolar component in the doublet splitting reflects directly the orientation of a particular bond vector with respect to an alignment tensor that is determined by the shape of the molecule. Knowing the structure of the protein this allows a back calculation of the RDCs and yields a set of theoretically predicted residual dipolar couplings[112]. Figure 2.7.1, depicts four residues extracted from a *J*-resolved 3D HNCA experiment that was recorded without scalar decoupling in the ^{13}C evolution period, see section 4.9.1.5 for experimental details. H^N-N and Hα-Cα residual dipolar couplings were fitted subsequently against the crystal structure of MalF-P2. A correlation of those back-calculated RDCs and the experimental RDCs are shown in Figure 2.7.2. RDC data of MalF-P2 in absence of MalE in solution render a fairly good fit for the two individual domains alone. For domain 1, a correlation coefficient of R = 0.78 and 0.81 for H^N-N and Hα-Cα, respectively, can be determined. Similarly, a correlation coefficient of R = 0.83 and R = 0.87 for H^N-N and Hα-Cα for domain 2 is obtained. A better fit of the correlation of the individual domains and the complete P2 loop of MalF might be achievable for a higher resolution X-ray structure. The X-ray structure of the maltose transporter[45] has a resolution of 2.8 Å rmsd at the well defined regions and 3 - 4 Å rmsd at regions far away from the lipid bilayer, which leaves some uncertainties concerning local structural fluctuations. No correlation is observed if all residues of the P2 loop of MalF are subjected to the fitting procedure (correlation coefficient R = 0.65 and R = 0.55 for H^N-N and Hα-Cα, respectively). The fit only improves marginally when the C-terminus (residues 260-275) containing the amphipathic helix is taken out of the fit. The correlation coefficient then improves; R = 0.68 and R = 0.58 for H^N-N and Hα-Cα, respectively. All correlation coefficients are summarized in Table 2.7.1. The fact that predicted and experimental RDCs match for the two domains indicates that the overall structure of the individual domains is preserved, but the relative orientation of the two domains of MalF-P2 is different in the crystal and in solution. This may be expected as the relative orientation of the two domains of MalF-P2 might change upon binding to MalE.

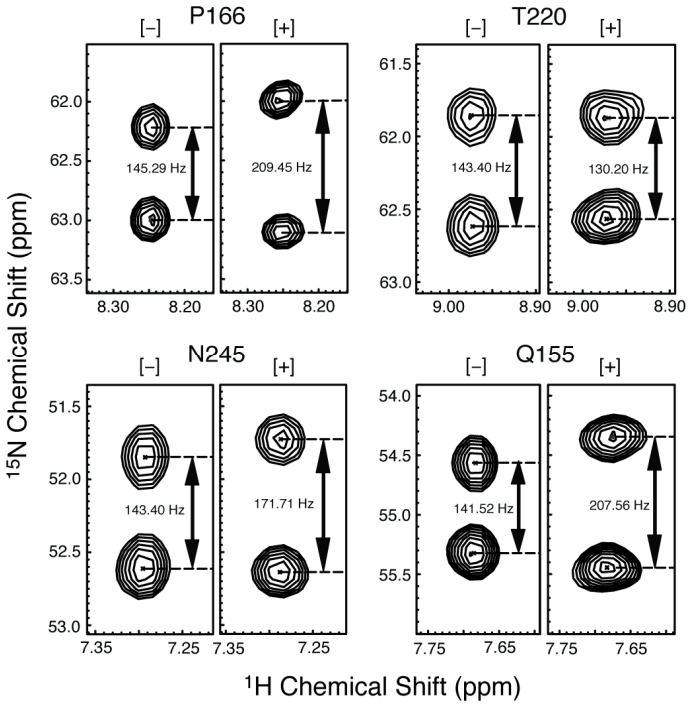

Figure 2.7.1. Hα-Cα residual dipolar couplings of MalF-P2 for selected residues.

Hα-Cα residual dipolar couplings (RDCs) for selected residues in MalF-P2. Pf1 phages (7mg/ml yielding a residual quadrupolar coupling of the D_2O resonance on the order of 17.5 Hz) were used as an alignment medium[113,114]. [-] and [+] indicate that the respective spectrum was recorded in the absence and presence of alignment medium. All spectra were recorded without maltose in the sample buffer.

Table 2.7.1. Correlation coefficients calculated by experimental RDC values for MalF-P2 and MalF-P2/MalE.

Correlation coefficients calculated by fitting experimental RDC values to the X-ray structure of $MalFGK_2$-E using the programs PALES[112]. High correlation coefficients indicate that the conformations in the crystal and in solution are comparable. Domain 1: residues 91-113 and 209-259; domain 2: residues 117-207; full length MalF-P2: 93-275, full length MalF-P2 without the C-terminus: 93-260.

Free MalF-P2		Dom 1	Dom 2	Dom 1 + Dom 2	Dom 1 + Dom 2 + C-term
H^N-H	all amino acids	0.78	0.87	0.68	0.65
	– crystal contacts	–	0.87	0.66	0.62
HαCα	all amino acids	0.81	0.83	0.58	0.55
	– crystal contacts	–	0.88	0.56	0.56
MalF-P2 + MalE					
H^N-H	all amino acids	0.90	0.52	0.64	0.64
	– crystal contacts	–	0.58	0.70	0.69

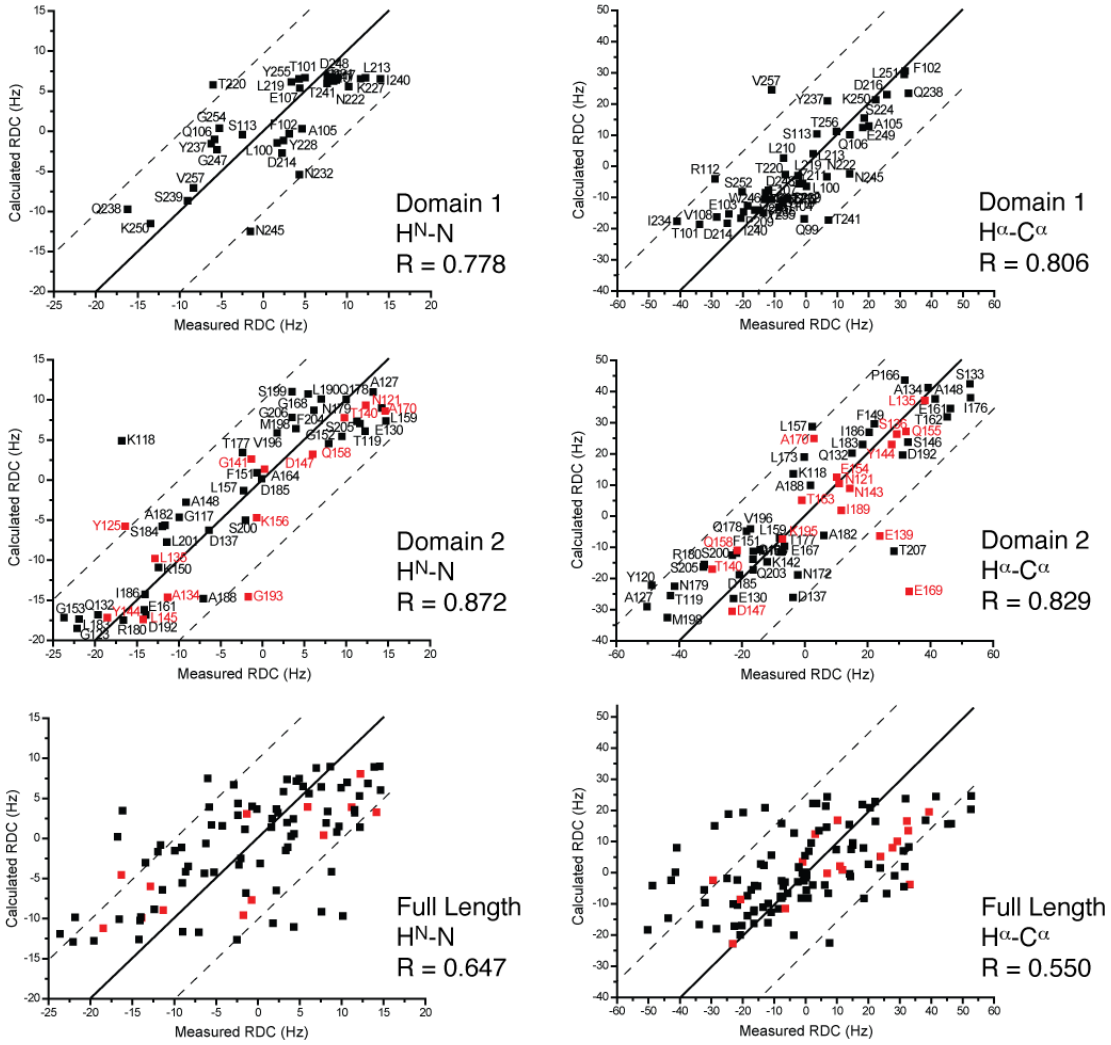

Figure 2.7.2. Calculated versus experimental RDC values, H^N-N and Hα-Cα, for MalF-P2.
Predicted RDC values versus experimental RDCs of MalF-P2 using the program PALES[112]. Using H^N-N and Cα-Hα RDC data, domain 1 (residues 91-113; 209-259) correlates with an average R-value of 0.79, domain 2 (residues 117-207) with an R-value of 0.85. All experimental RDC values could be fit to the full structure of the MalF-P2 loop yielding an average R-value of 0.59. Residues, which are colour coded in red are involved in crystal contacts to a MalK molecule in the symmetry related unit cell.

In order to probe if MalF-P2 is structurally reorganized upon binding to MalE, the RDC experiments were repeated in the presence of MalE. Figure 2.7.3 shows a correlation between experimental and theoretical RDC couplings for MalF-P2 in presence of MalE. The theoretical RDC values are again obtained by back-calculation from the X-ray structure of the maltose transporter using the program PALES[112]. The correlation coefficient for domain 1 remains high, indicating that its structure is not perturbed upon addition of MalE. On the other hand, the correlation coefficient for domain 2 decreases significantly. This implies that the structure of domain 2 is different to the structure of domain 2 in the crystal. The overall correlation coefficient remains small (R = 0.64 for all amino acids, and R = 0.64 for full-length MalF-P2 without the C-terminal residues, respectively), which could be interpreted as that the structure of domain 2 are different in solution and the crystal structure, but its overall

orientation to domain 1 is the same. Residues, which are colour coded in red (domain 2), are involved in crystal contacts to a MalK molecule in the symmetry related unit cell in the X-ray structure.

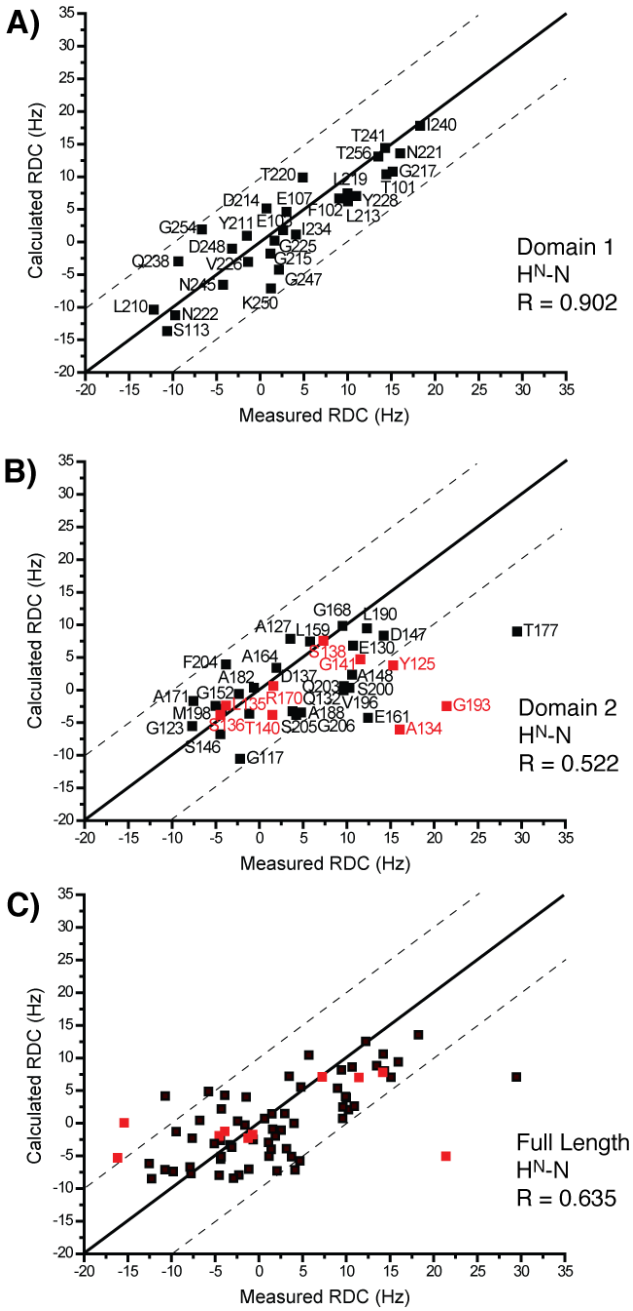

Figure 2.7.3. Calculated versus experimental RDC values for MalF-P2 in complex with MalE.
Experimental versus predicted RDC values for MalF-P2 in presence of MalE. The theoretical RDC values are calculated from the X-ray structure using the program PALES[112]. A) RDC data fitted to domain 1 (R = 0.91), domain 2 (R = 0.522) and to full-length MalF-P2 (R=0.64) are represented in A, B and C, respectively. A lower correlation coefficient in the presence of MalE suggests that the conformation of domain 2 of MalF-P2 in the crystal structure and in solution is different. Residues, which are colour coded in red, are involved in crystal contacts to a MalK molecule in the symmetry related unit cell.

A better structural fit of the MalF-P2 and MalE complex could possibly be achieved by addition of another set of RDCs for the bound state. The large protein complex was however not accessible by *J*-resolved 3D HNCA experiment used for the free MalF-P2. Addition of ^{2}H MalF-P2 to MalE was not enough to give favourable relaxation times to retrieve any 3D experimental data for the protein complex.
A structure refinement calculation with Xplor[115] of bound MalF-P2 to MalE with H^N-N RDCs added was conducted. Different weighting of the H^N-N RDCs produced different final structures with low convergence. Single sets of RDCs have a large degeneracy (16-fold)[116], within the experimental error, giving high discrepancy in the structure calculation[117]. Multiple sets of RDCs (i.e. Hα-Cα, Cα-CO, Hx-Cx) or the use of multiple alignment media can greatly reduce the degeneracy. H^N-N RDC fits for bound MalF-P2 to Xplor[115] refined structures yielded R-values between 0.5 – 0.7.

2.8 Molecular dynamics experiments of the MalF-P2 and MalE interaction

Free MD simulation can provide dynamic structural data as observed for macromolecules in solution. This information can then be used to fit experimentally determined data, i.e. RDCs. Structure refinements by computational methods with one set of RDCs (section 2.7) yielded poor convergence and MD simulations were carried out to retrieve an average structure of MalF-P2 in complex with MalE to improve structural data.
A GROMACS molecular dynamics simulation[118] was performed for the MalF-P2 – MalE complex *in silico*. The crystal structures of MalE and MalF-P2 in the $MalFGK_2$ complex were used as starting points for the simulated annealing. The full length sequence of MalE were calculated for each simulation ensemble. The C-terminal part, residues: 266 – 275 of MalF, was discarded as earlier data shows no interaction between MalE and MalF-P2 in this region[45], and that the C-terminal α-helix α3 cannot fold up independently in solution. The molecular system was placed into a triclinic box and H_2O (at 310 K), 0.1 mM Na^+ and Cl^- ions were added to the system to represent physiological conditions. An initial waterbox refinement of 2.5 ns was used to equilibrate the system before simulated annealing. After equilibration over a period of 500 ps using a positional restraint of 1000 kJmol-1nm-2 on the backbone of the molecular system, a free simulation was performed for a period of 50 ns. During the course of the simulation, the trajectory was sampled every 5 ps. Rmsd trajectories were plotted for MalE, MalF-P2 and the MalF-P2 – MalE complex, aligned to the MalE starting structure (Figure 2.8.1, panel A). The MalE trajectory shows that MalE maintains its initial structure and any dynamic behaviour is only seen for surface exposed loops in the C-terminal lobe of MalE. MalE behaves as a rigid body with rmsd amplitude of 3 Å throughout the simulation. In contrast to MalE, MalF-P2 has large dynamic variations throughout the simulation (Figure 2.8.1). The rmsd of the trajectory shows an initial rise at the starting point of the simulation. Large initial movements to accommodate the binding surface to MalE properly, cause this initial "jump" in the rmsd. After 25 ns a local energy minimum, conformation equilibrium, can be seen (Figure 2.8.1 panel A) and MalF-P2 has positioned itself onto the binding surface of MalE. The second half of the simulation shows a higher rmsd trace, as the two domains of MalF-P2 are dynamic in their interaction with MalE.

Analysis of the dynamics of the two individual domains of MalF-P2 gives an appreciation of their respective modes of movement towards the MalE binding surface. From NMR interaction studies, the major part of the interaction of MalF-P2 and MalE involves in the two α-helices α1 and α2 (see section 2.5.3). These α-helices are located in each of the two domains of MalF-P2. Both domains of MalF-P2 are reorientaing to lower their energy during the simulated annealing. Up to 25 ns of the calculation, the rmsd trace shows an average of 4.5 Å (Figure 2.8.2, panel A), and the domains domains 1 and 2 have a 7.5 Å and 6.0 Å rmsd, respectively (Figure 2.8.2, panel A). Comparison of the frames from the first 20 ns (Figure 2.8.2, panel B and C, blue cartoon structures in ensemble) to the frames after 40 and 50 ns (Figure 2.8.2, panel B and C, red cartoon structures in ensemble), there is a clear difference of the orientation of α-helix α1 in domain 1 forcing it to move towards domain 2 of MalF-P2 on the binding surface of MalE. The angle between the α-helix at 0 ns and 25 ns is on the order of 30^0. α-helix α2 in domain 2 also undergoes reorientations, but to a much less significant extent, roughly 10^0. The dynamics simulation after 30 ns shows the two domains reorienting independently at their respective binding interfaces. The two domains experience different orientations, which are almost perpendicular to each other (Figure 2.8.1 panel C and 2.8.2, panels B and C). The mass centre of α-helices α1 and α2 are set as mass centre for the domain movements in domain 1 and 2.

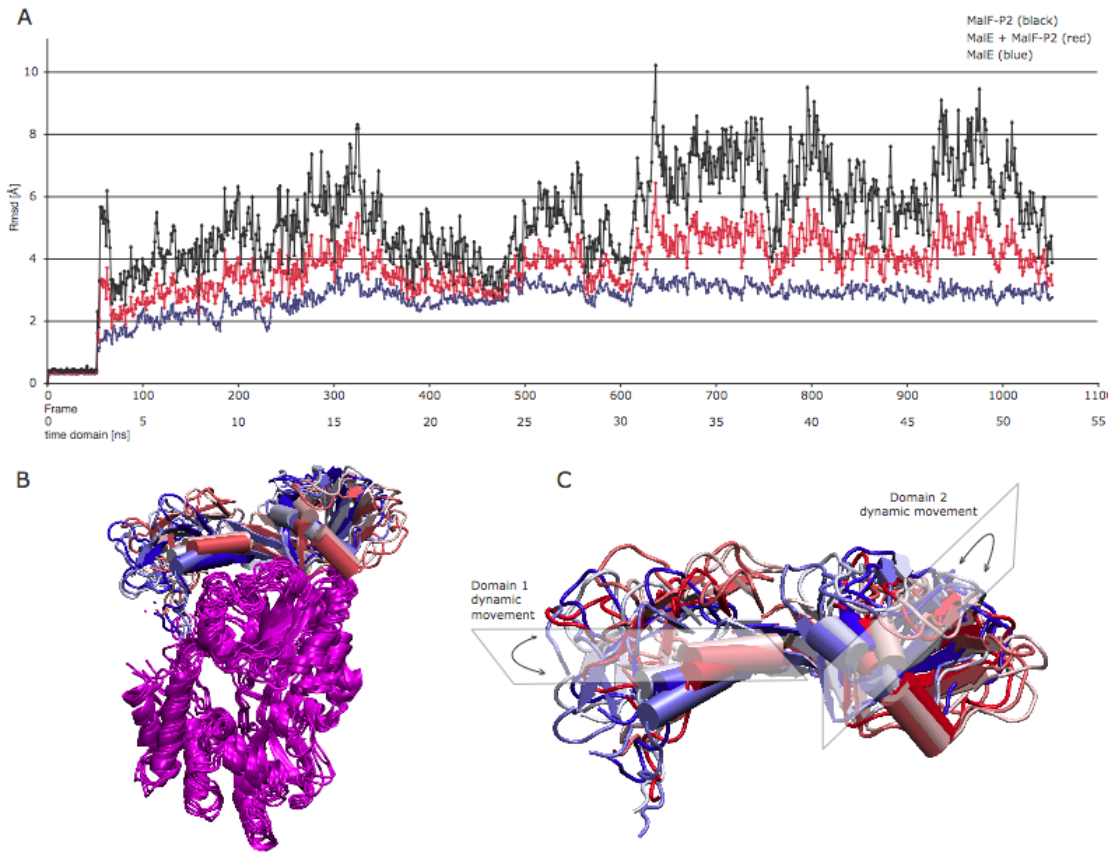

Figure 2.8.1. MD trajectories for the MalF-P2 complex.
Trajectory of rmsd for MD simulation of the MalF-P2 – MalE complex. Panel A shows the rmsd trajectory for the 50 ns dynamics simulation aligned to the MalE structure. The first 2.5 ns comprised a waterbox refinement before the simulation-annealing step. The rmsd trajectory of MalE alone is

represented in blue, MalF-P2 alone in black and the overall motion of the MalF-P2 – MalE complex in red in panel A. MalE keeps a rigid structure with low dynamics, backbone trace of full simulation, in violet, is represented in panel B. Panel B and C represent backbone traces for MalF-P2 (traces ranging from blue: 0 ns to red: 50 ns in panels B and C) and MalE (represented in violet in panel B). The modes of motion for the two domains of MalF-P2 are represented in panel B as transparent planes with arrows showing the direction of movement. MalF-P2 undergoes large restructuring in the initial step, before finding its alignment to MalE. A much larger dynamic behaviour can be seen for MalF-P2 over the full dynamics simulation, rmsd ranging from 4-9 Å in contrast to MalE that has a stable rmsd of 3 Å. Representations of MalF-P2 at different time steps of the simulation are drawn in panels B and C. Structure are shown for every 10 ns of the simulation with colours going from blue to white to red. The two domains can be seen having two different modes of motion on the structure of MalE. The two different modes are approximately perpendicular to each other with an amplitude of roughly 30^0 for domain 1 and 10^0 for domain 2, and are represented in panel C as transparent planes with α-helices α1 and α2 mass centre's as weight points for the motion. The two α-helices of MalF-P2, α1 and α2, make up the main interaction with MalE, accommodate large rearrangements of their positioning and still maintain the interaction to MalE. The large dynamics of the loose loops of MalF-P2 on the reverse side of the MalF-P2–MalE interface also influences the positioning of α1 and α2, making the interaction very complex. Figures in panels B and C were realised with VMD[119].

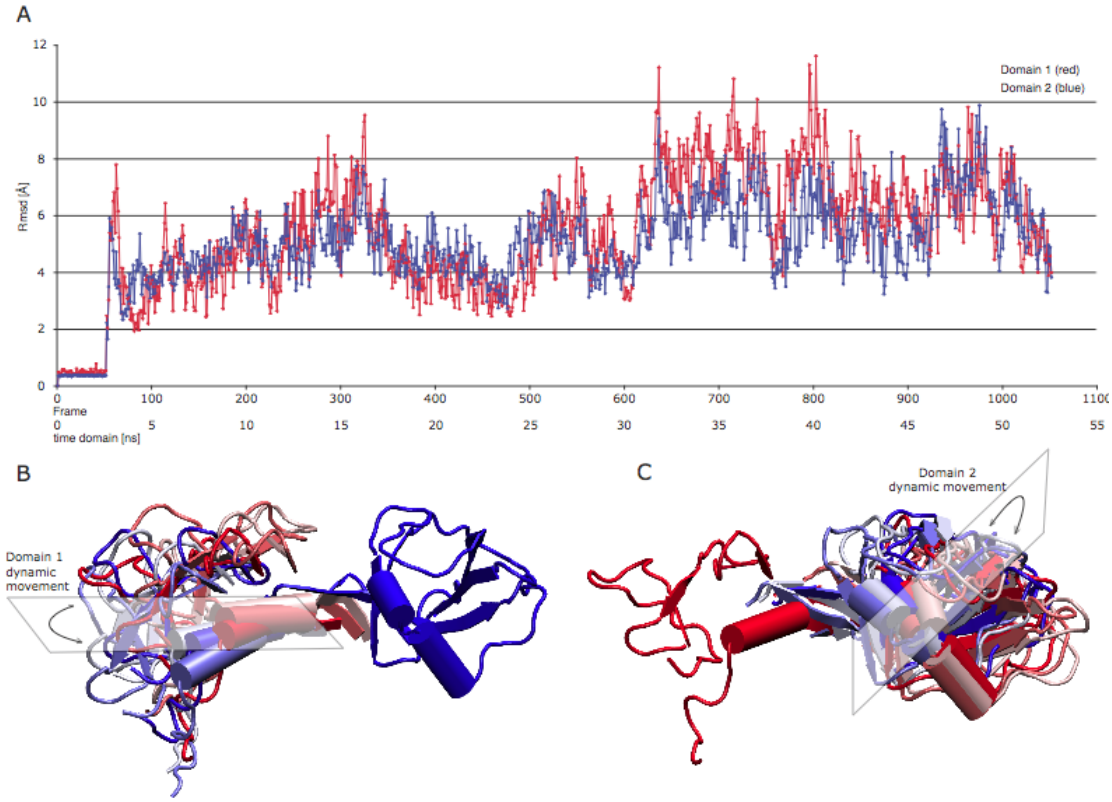

Figure 2.8.2. MD trajectories for domain 1 and 2 of MalF-P2.
Rmsd trajectories for the two domains of MalF-P2 for MD simulation of the MalF-P2 – MalE complex. The first 2.5 ns comprised a waterbox refinement for the full system before the simulation-annealing step. The rmsd trajectories of domain 1 and 2 are represented in red and blue respectively in panel A. Panel B and C represent the simulated annealing for domain 1 (B) and domain 2 (C), keeping the opposite domain fixed at 0 ns, together with their direction of motion towards the interface of the MalF-P2 – MalE complex. Backbone traces in panel B and C for structures calculated each 10 ns are represented from blue to white to red. The rmsd trajectories equilibrate after the initial energy minimization and can be seen ranging from 4-9 Å. The two domains of MalF-P2 both have large amplitudes of motion towards the MalE - MalF-P2 interaction surface. The two different modes of motion are approximately perpendicular to each other with an amplitude of roughly 30^0 for domain 1

and 10^0 for domain 2, and are represented in panel B and C as transparent planes for respective domain. α-helices α1 and α2 were used as weight points for the dynamic motions as to their high relevance to the binding. Figures in panels B and C were realised with VMD[119].

A second 50 ns molecular dynamics simulation was performed to monitor the rmsd traces from the respective calculations. The two calculations converged and showed similar dependence on the dynamic behaviour of the MalF-P2 – MalE interaction.

These results indicate that the interaction of MalF-P2 and MalE is highly dynamic, which is supported by NMR RDC data, and allows different modes of binding of MalF-P2 to MalE when present in the full $MalFGK_2$ complex.

2.9 Relaxation experiments of MalF-P2 in absence and presence of MalE

Solution NMR dynamic studies on the μs – ms timescale were performed on MalF-P2 and the MalF-P2 – MalE complex in solution. These experiments give valuable information regarding domain movements during binding events and any dynamic properties of specific regions. T_1- and T_2-relaxation experiments clearly show the dynamics of the hinge between the two domains of MalF-P2 and its ability to rearrange its domain conformation to fully access the binding interface of MalE. Further on, the exchange rate, k_{ex}, of the complex formation between MalF-P2 and MalE could be determined to ~0.6 ms in absence and presence of maltose. Together with the ITC data this shows that the interaction between MalF-P2 and MalE is not connected to substrate availability of MalE.

2.9.1 ^{15}N T_1-experiments and ^{15}N T_2-experiments

T_1 and T_2 relaxation times were measured for the individual residues of MalF-P2 and are shown in Figure 2.9.1.1. Residues 114-117 in the linking region between domain 1 and domain 2 can have a lower than average T_1 relaxation and higher than average T_2 relaxation. Its counterpart at residue 208-209 shows no significant difference to the average relaxation. This indicates that only the N-terminal part of the hinge region is really flexible and the C-terminal part of the hinge is rigid, most likely to be too short to accommodate any larger structural changes. The C-terminal region, residue 260-270, where the amphipathic α-helix is found in the crystal structure of $MalFGK_2$-E[45], shows lower than average T_1 relaxation times, indicating that the helix cannot properly fold in solution. Also, the region before the second α-helix, residues 162-170, displays a lower than average T_1 relaxation and higher than average T_2 relaxation. In the crystal structure, this α-helix has a kink and the NMR angle restraints (Talos predictions) and NOE data show that this is a straight extended α-helix in solution. This mobility might be needed to accommodate the kink in the α-helix when bound to MalE. Residues 270-274 also shows a much higher than average T_2 relaxation rate, but with no significant difference in T_1 relaxation, indicating that the C-terminal region is structured but flexible towards the rest of the protein.

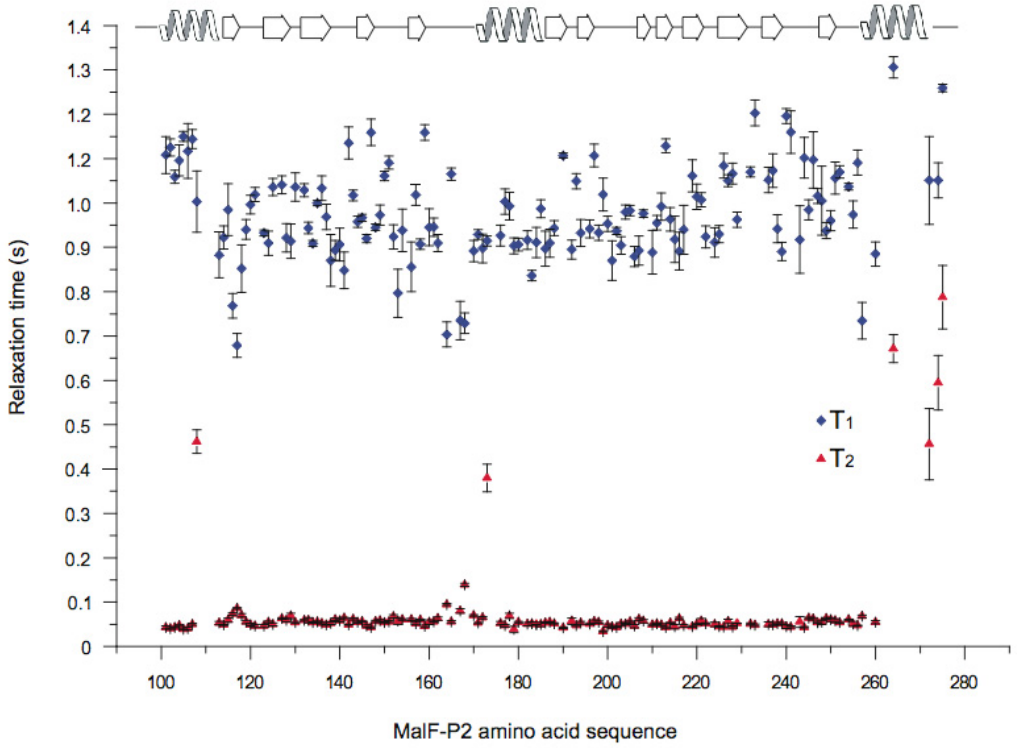

Figure 2.9.1.1. T_1- and T_2-relaxation rates of MalF-P2.
Relaxation properties of the backbone amides of MalF-P2. T_1 (blue) and T_2 (red) relaxation times of all assigned amino acids are shown, with secondary structural elements MalF-P2 from the crystal structure of $MalFGK_2$[45].

2.9.2 ^{1}H-^{15}N heteronuclear NOE measurements

^{1}H-^{15}N heteronuclear NOE measurements were performed for MalF-P2 (500 µM) and for MalF-P2 in the presence of equimolar ratio of MalE (1 mM). In agreement with T_1 and T_2 relaxation data in section 2.9.1, the C-terminal part of the protein is highly flexible, as judged by the low (negative) NOE values (Figure 2.9.2.1 in blue). This is also seen for the complex of MalF-P2 and MalE (Figure 2.9.2.1 in red). Interestingly, the heteronuclear NOE data show the same behaviour indicating a fast exchanging complex. The heteronuclear NOE data also show the same flexibility in the region at the second α-helix as observed in the T_1 and T_2 relaxation data, residues 160-170.

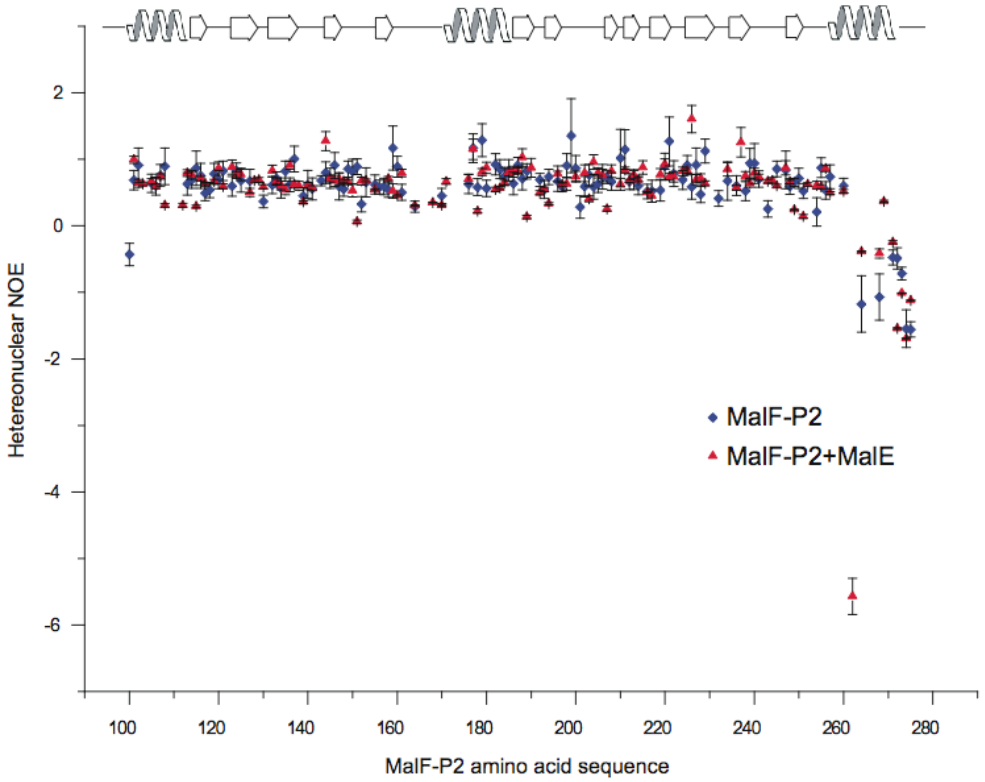

Figure 2.9.2.1. ^{1}H-^{15}N heteronuclear NOE relaxation of MalF-P2.
^{1}H-^{15}N heteronuclear NOE experiments for MalF-P2 (blue) and for the MalF-P2 and MalE complex (red) at equimolar concentrations.

2.9.3 Molecular weight determination from relaxation rates

From T_1 and T_2 relaxation data the correlation time, τ_c, can be calculated for the full-length protein MalF-P2 (residues 91-274) and for the two domains (domain 1: residues 91-113; 209-259, and domain 2: residues 117-207) as derived from the secondary structure data (section 2.4) and the crystal structure of the maltose transporter[45]. According to methods of Gryk *et al.*[120] a τ_c of 8.4 ns for full-length MalF-P2, 8.9 ns for domain 1 and 9.9 ns for domain 2 could be determined. This correlation time fits to the size of a globular molecule with a size of 13.9, 14.8 and 15.8 kDa for the full length, domain 1 and domain 2 respectively (Figure 2.9.3.1). The smaller than expected molecular weight is in agreement with measured line widths from ^{1}H-^{15}N and ^{1}H-^{13}C heteronuclear NMR experiments, where narrower line widths than expected are observed. Proteins with internal domain motions have both an overall tumbling rate, but also internal dynamics that causes the macromolecule to be more dynamic than globular proteins. The relaxation data show that the two domains of MalF-P2 tumble independently from each other. It must be emphasized that this method of estimating the overall tumbling rate τ_c, is very coarse and gives a rough estimate.

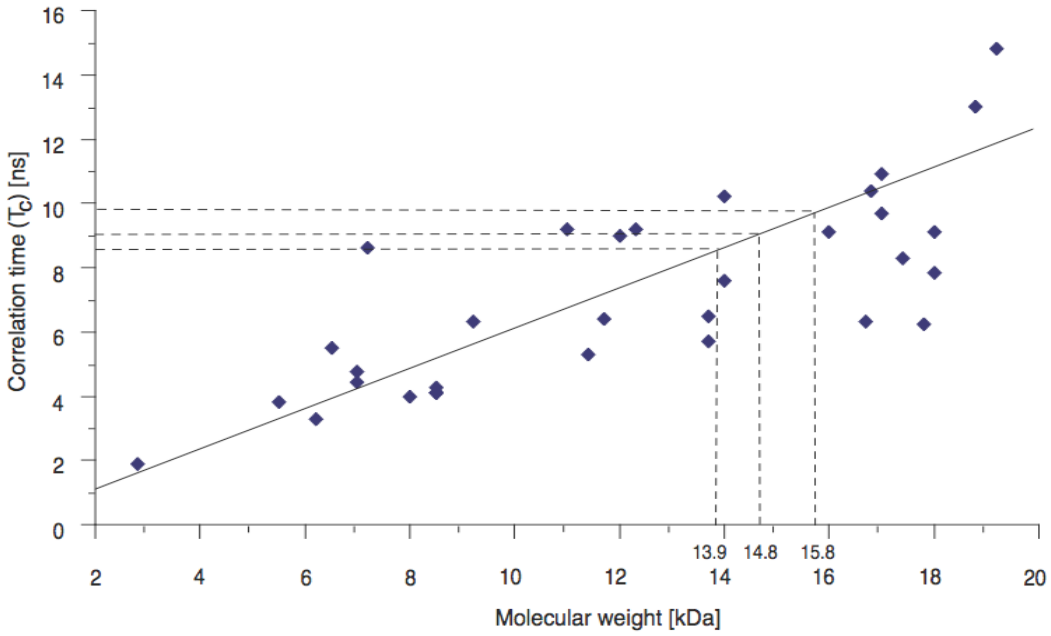

Figure 2.9.3.1. Overall tumbling rate and predicted molecular weights of MalF-P2 domains 1 and 2. Overall tumbling rates τ_c and their molecular weights for a range of proteins of different sizes and shapes[120]. The full-length protein MalF-P2 was found to have an overall tumbling rate, τ_c, of 8.4 ns, the domains 1 and 2, 8.9 and 9.9 ns, respectively. These values were fitted to the overall tumbling, τ_c, of a range of proteins. This yields an estimate of the molecular weights. Full-length MalF-P2 fits to a 14 kDa globular protein and the two domains 1 and 2 fit to 15 and 16 kDa globular proteins respectively.

2.9.4 Carr-Purcell Meiboom-Gill (CPMG)

NMR CPMG experiments were conducted to probe the molecular dynamics of MalF-P2 and of MalF-P2 in complex with MalE on the μs – ms timescale[121]. This timescale provides information on conformational changes in ligand binding and folding – unfolding events[121-123]. The chemical exchange implied in these processes contributes to the transverse relaxation rate in the laboratory frame (R_2). A general expression for the phenomenological transverse relaxation rate constant for site A, $R_2(1/\tau_{cp})$, is given by[124],

$$R_2(\tau_{cp}) = R_2^0 + \frac{1}{2}\left(k_{ex} - \frac{1}{\tau_{cp}} \cosh^{-1}\left[D_+ \cosh(\eta_+) - D_- \cos(\eta_-) \right] \right) \qquad [2.9.4.1]$$

where

$$D_\pm = \frac{1}{2}\left[\pm 1 + \frac{\psi + 2\Delta\omega^2}{\left(\psi^2 + \zeta^2\right)^{1/2}} \right]$$

$$\eta_{\pm} = \frac{\tau_{cp}}{\sqrt{2}}\left[\pm\psi + \left(\psi^2 + \zeta^2\right)^{1/2}\right]^{1/2}$$

$$\psi = \left(p_a k_{ex} + p_b k_{ex}\right)^2 - \Delta\omega^2 + 4 p_a p_b k_{ex}^2$$

and

$$\zeta = 2\Delta\omega\left(p_a k_{ex} + p_b k_{ex}\right)$$

in which τ_{cp} is the delay between 180^0 pulses in the CPMG pulse train[125,126], $R_2{}^0 = (R_{ex} + R_2)/2$, p_a and p_b the populations of the two states, k_{ex} the exchange rate and $\Delta\omega$ is the frequency difference between R_{ex} and R_2. The chemical exchange contribution to transverse relaxation, R_{ex}, is defined as the difference between the apparent relaxation rate constants in the slow and fast pulsing limits[127]. The limiting values of R_{ex} in slow and fast exchange are

$$R_{ex} = p_b k_{ex} \qquad \left(k_{ex} / \Delta\omega \to 0\right) \qquad [2.9.4.2]$$

$$R_{ex} = p_a p_b \Delta\omega^2 / k_{ex} \qquad \left(k_{ex} / \Delta\omega \to \infty\right) \qquad [2.9.4.3]$$

respectively. For very fast exchange R_{ex} depends quadratically on the static field strength[128,128]. For very slow exchange R_{ex} is independent of the static field strength[130]. The functional form of R_{ex} given by the master equation (Eq. [2.9.4.1]) is not overly sensitive to the values of R_a and R_b provided

$$\left|R_a - R_b\right| / \Delta\omega = \left(p_a - p_b\right) \qquad [2.9.4.4]$$

It is convenient to introduce a proportionality constant or scaling factor. This scaling factor is denoted α and can be written as:

$$\alpha = \frac{2\left(k_{ex} / \Delta\omega^2\right)}{1 + \left(k_{ex} / \Delta\omega\right)^2} \qquad [2.9.4.5]$$

provided that $p_a \gg p_b$, α satisfies the condition $0 \le \alpha \le 2$, which generalizes the limiting cases in Eqs. [2.9.4.2-3].
α is a function of $k_{ex}/\Delta\omega$ and therefore defines the NMR chemical shift timescale as:

$0 \le \alpha < 1$	slow exchange
$\alpha = 1$	intermediate exchange
$1 < \alpha \le 2$	fast exchange

For MalF-P2 alone a two-site exchange system is considered for the conformational changes of the two domains. In the case of the MalF-P2–MalE complex, the

interaction is treated as a two-site exchange system as well. Solving a second-order reaction of two first-order reactants[131] with the ITC determined K_D (25 μM) yields the concentrations of the free MalF-P2 and the MalF-P2 – MalE complex. At equimolar concentrations of 700 μM for MalF-P2 and MalE, 99% is involved in complex formation. The low population of conformational changed free MalF-P2 is therefore negligible.

The ^{15}N $R_2(1/\tau_{cp})$ measurements were performed at two static magnetic fields, 600 and 900 MHz ^1H Larmor frequency, for MalF-P2 and at one static field for the MalF-P2 – MalE complex, 750 MHz ^1H Larmor frequency, at 300 K. In-phase and anti-phase T_2 relaxation was averaged within the same experiment, in order to eliminate any τ_{cp}-dependent affect arising from the differential relaxation[132]. ^1H 180^0 pulses were applied to eliminate ^1H-^{15}N dipolar/^{15}N chemical shift anisotropy relaxation interference[133].

$1/\tau_{cp}$ dependent data obtained were fit using an in-house python script (written by Sam Asami, FMP Berlin) based on the methods of the groups of Palmer[132] and Loria[125]. The fitting algorithm uses a 20,000 population with stochastic values for R_2^0, p_a, k_{ex}, and $\Delta\omega$ in the ranges described by Kovrigin *et al.*[124] A *fitness* is calculated for each parameter set for whole population by:

$$\theta_i = \frac{1}{rmsd} \rightarrow \infty \qquad [2.9.4.6]$$

$$fitness = \frac{\phi_i}{\overline{\phi_i}} \qquad [2.9.4.7]$$

An intermediate population is created. Based on the *fitness* the likelihood for keeping a specific parameter set from the initial population, for the next generation, is defined. The *fitness* determines the chance for values being kept for the creation of the intermediate populations. Values for the next generation are created by recombination of parameter sets in the intermediate populations. The recombination of values from the different parameter sets is stochastic. High *fitness* thus equals to a higher chance of preserving the values for the next generation. 10 generations were performed for the full computation.
To obtain the intrinsic error of the different parameters, a complex algorithm is needed and was not used in this work. Curve uncertainties were determined by the rmsd of the best-fit to the raw data:

$$rmsd = \sqrt{\frac{1}{n}\sum\left(\rho^{obs} - \rho^{fit}\right)^2} \qquad [2.9.4.8]$$

All data points were weighted equally with k_{ex}^{-1} extracted from the curves (Table 2.9.4.1-2). Values of α were calculated from data recorded at two fields for MalF-P2

and for data recorded at one field for the MalF-P2 - MalE complex (Tables 2.9.4.1-2). Figures 2.9.4.1-2 show the best-fit for selected residues for MalF-P2 and the MalF-P2 – MalE complex.

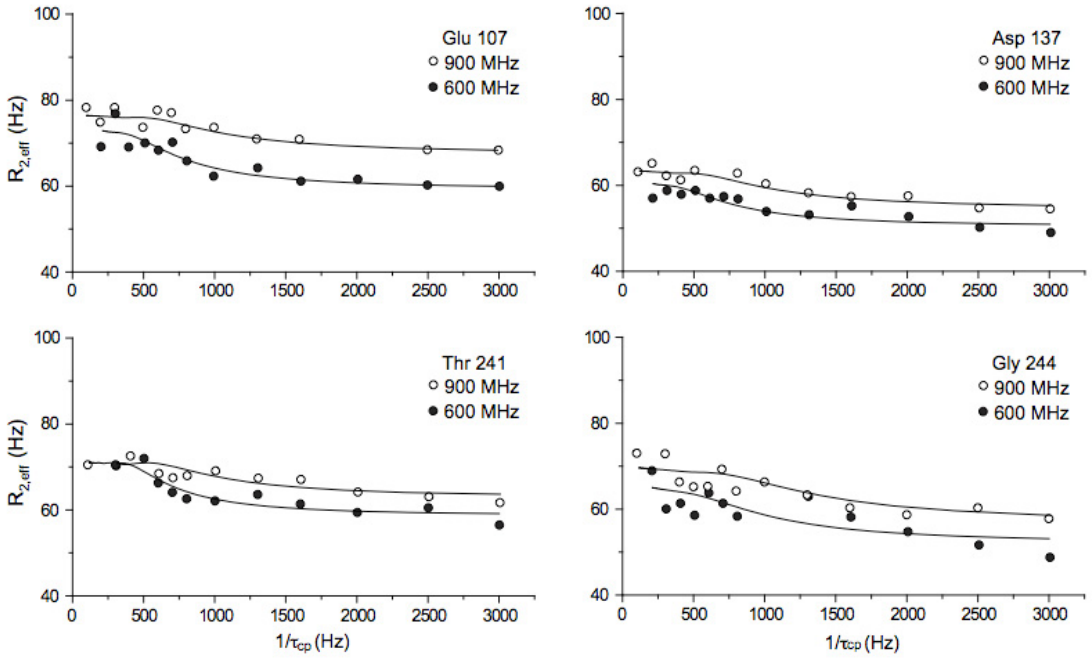

Figure. 2.9.4.1. Relaxation dispersion curves of selected residues for MalF-P2.
Effective transverse ^{15}N relaxation rates, R_2^{eff}, as a function of the CPMG field strength, $1/\tau_{cp}$, for Met107, Phe145, Asp152 and Ile182 of MalF-P2. The best-fit dispersion curves fitted as described by Kovrigin *et al.*[124] are shown (solid lines). Data recorded at 900 and 600 MHz ^1H Larmor frequency are indicated as open-circles and closed-circles, respectively. Values of k_{ex}^{-1} for these residues are [ms]: 1.06 for Met107, 0.697 for Phe145, 0.511 for Asp152, 0.802 for Ile182. All residues depicted have a p_a = 0.99 from the best-fit.

Table 2.9.4.1. Relaxation dispersion curve constants for MalF-P2.
K_{ex}^{-1} (ms), for the conformational changes were extracted from the best-fit curves. The population, $p_{a,ave}$, were directly extracted from the fit together with the $rmsd_{ave}$ (Hz) of the best-fit curves to the average fit. The scaling factor α_{ave} was calculated from k_{ex} and $\Delta\omega_{ave}$ according to Eq. [2.9.4.5]. Values for α_{ave}, $p_{a,ave}$ and $rmsd_{ave}$ are shown as averages of best-fit curves for the simultaneous fit.

Residue	k_{ex}^{-1} (ms)	$p_{a, ave}$	$rmsd_{ave}$ (Hz)	α_{ave}	$\Delta\omega_{600}$ (Hz)	$\Delta\omega_{900}$ (Hz)
106	0.513	0.99	2.49	2.0E-04	3172	4759
107	1.06	0.99	2.02	2.6E-04	2007	3011
127	0.692	0.99	1.76	3.0E-04	2220	3330
135	0.698	0.90	1.53	1.3E-03	354.9	532.3
137	0.962	0.99	1.65	3.0E-04	1920	2880
141	0.524	0.99	1.75	2.2E-04	2978	4468
144	7.35	0.91	1.67	1.1E-04	1241	1862
145	0.697	0.99	1.18	2.2E-04	2663	3994
150	0.633	0.99	1.30	5.1E-04	1529	2294
152	0.511	0.99	1.32	2.0E-04	3152	4729
158	8.18	0.91	1.09	1.0E-04	1277	1915
161	0.642	0.81	1.94	1.2E-03	332.1	498.2
180	0.560	0.99	1.41	4.9E-04	1620	2431
182	0.802	0.99	1.09	3.5E-04	1872	2809
185	0.519	0.94	1.61	9.6E-04	450.4	675.7
204	0.781	0.99	2.15	2.7E-04	2249	3374
211	8.80	0.88	2.10	1.0E-04	1192	1788
213	0.501	0.99	3.02	2.9E-04	2492	3739
217	1.90	0.97	3.15	3.1E-04	1424	2136
220	0.676	0.97	2.21	1.0E-03	676.6	1014
241	6.15	0.93	1.55	4.6E-05	2130	3195
244	0.627	0.99	2.85	2.7E-04	2429	3642
248	1.88	0.98	1.21	1.1E-04	2409	3612

An average exchange rate of residues with a p_a of > 0.98 gives a conformational sampling with k_{ex}^{-1} of 0.738 ms. From the bulk determined p_a values, MalF-P2 comprise a 1% population with conformational differences. The exchange between the two conformations are evaluated by the scaling factors, α, of the best-fit of the dispersion curves (Eq. [2.9.4.5] and Table 2.9.4.1) and are determined to be in the slow exchange regime. Chemical shift changes of ^{15}N range between 0.531 – 5.32 ppm as extracted from the $\Delta\omega$ of the relaxation dispersion curves (Table 2.9.4.1). The two conformers, $p_a = 0.99$ and $p_b = 0.01$, exchange with the bulk exchange rate, k_{ex}^{-1} = 0.738 ms. Together with the calculated τ_c time (section 2.9.3), the data can be interpreted as that only a small population of MalF-P2, 1%, encompasses a different conformation. Also, the two domains undergo motion independent of each other explaining the lower that anticipated τ_c values for a protein of 21 kDa (section 2.9.3). Larger k_{ex}^{-1} values can be seen for residues: Tyr144 (7.35 ms), Gln158 (8.18 ms), Tyr211 (8.80 ms). Tyr144 and Gln158 are located in loop regions, one can speculate

that the loops undergo slower dynamics to the more rigid β-sheets, β4 and β5. Causing the loops to exchange between two states. Tyr211 is in close vicinity to α1; the residue undergoes conformational changes slower then the bulk rate. This exchange could be responsible for keeping domain 1 more tightly packed for MalF-P2 in solution.

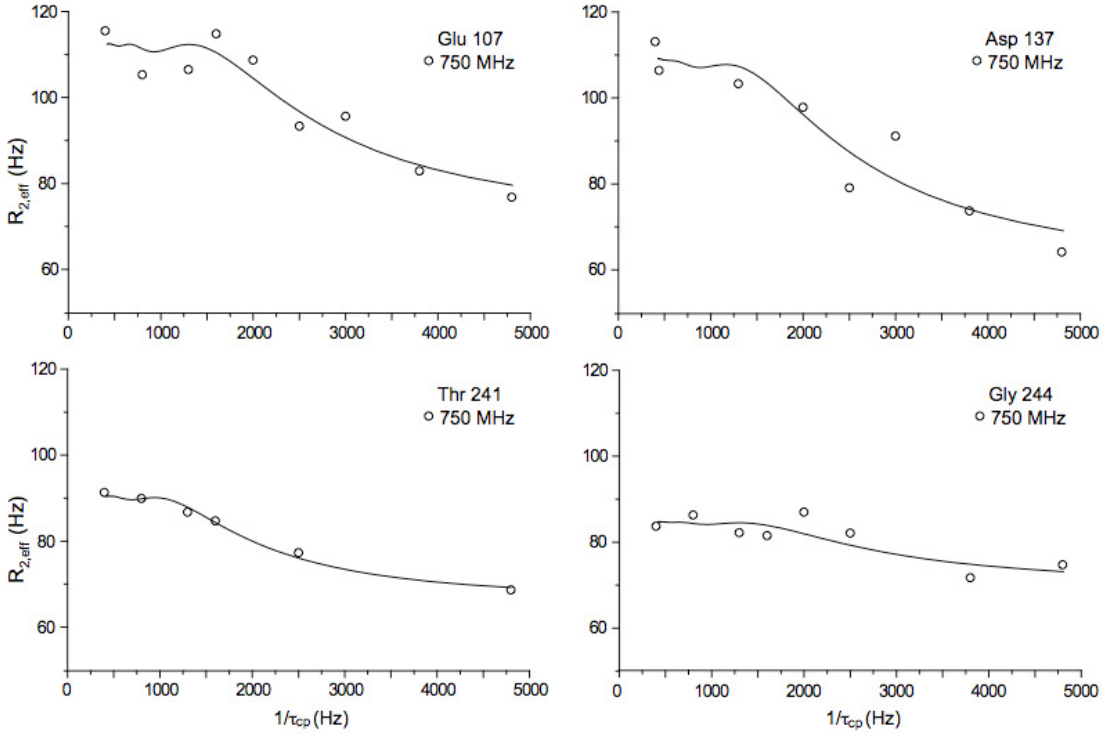

Figure. 2.9.4.2. Relaxation dispersion curves of selected residues for the MalF-P2 – MalE complex. Effective transverse ^{15}N relaxation rates, R_2^{eff}, as a function of the CPMG field strength, $1/\tau_{cp}$, for Ser114, Tyr137, Thr240 and Ala241 in the MalF-P2 – MalE complex. The best-fit dispersion curves fitted as described by Kovrigin et. al[124] are shown (solid lines). Data recorded at 750 MHz ^1H Larmor frequency are indicated as open-circles. Values of k_{ex}^{-1} for these residues are [ms]: 0.534 for Ser114, 0.518 for Tyr137, 0.559 for Thr240, 0.722 for Ala241. All residues depicted have a $p_a = 0.97$ from the best-fit.

Table 2.9.4.2. Relaxation dispersion curves for the MalF-P2 – MalE complex.
Relaxation dispersion curve constants for the MalF-P2 – MalE complex. K_{ex}^{-1} (ms), for the conformational changes were extracted from the best-fit curves. The population, p_a, were directly extracted from the fit together with the rmsd (Hz) of the best-fit curves to the average fit. The scaling factor α were calculated from k_{ex} and $\Delta\omega$ according to Eq. [2.9.4.5].

Residue	k_{ex}^{-1} (ms)	p_a	rmsd (Hz)	α	$\Delta\omega_{750}$ (Hz)
106	0.504	0.98	5.42	3.8E-04	2530
107	0.106	0.95	7.98	2.8E-05	8142
114	0.534	0.98	3.81	5.6E-05	7931
127	0.658	0.98	6.27	4.4E-05	8168
135	0.551	0.96	8.59	8.6E-05	6250
137	0.518	0.97	5.46	6.7E-05	7316
141	0.746	0.98	5.21	4.0E-05	8082
144	0.560	0.98	11.5	5.2E-05	8101
145	0.793	0.98	9.69	3.7E-05	8106
150	0.540	0.97	6.91	5.7E-05	7860
152	0.511	0.98	3.60	5.9E-05	7906
158	0.505	0.98	3.00	7.1E-05	7185
161	0.647	0.97	3.44	4.5E-05	8189
180	4.53	0.80	8.33	2.5E-05	4193
182	0.538	0.97	2.84	8.1E-05	6514
185	0.511	0.98	2.33	5.9E-05	7939
204	0.616	0.97	5.95	4.7E-05	8158
211	0.693	0.98	4.02	4.2E-05	8172
213	3.35	0.87	8.85	3.0E-05	4445
217	4.29	0.84	10.4	2.9E-05	3993
220	1.21	0.96	8.45	2.4E-05	8176
240	0.559	0.97	3.07	1.6E-04	4437
241	0.722	0.98	0.817	7.4E-05	5994
244	0.616	0.99	2.80	4.7E-05	8111
248	0.506	0.99	1.57	1.7E-04	4444

An average exchange rate of residues with a p_a of > 0.96 gives a k_{ex}^{-1} of 0.597 ms. 96% of MalF-P2 is in complex with MalE. With the error estimation of the best-fit curves this is in agreement with ITC data[131] as described earlier in this section. From the exchange rate the interaction time of MalF-P2 in complex with MalE is determined. The interaction can thus be seen as a slow exchanging system with the scaling factors, α, being on the order of $1 \cdot 10^{-4}$ s (Eq. [2.9.4.5] and Table 2.9.4.2). Chemical shift changes of ^{15}N range between 3.46 – 10.9 ppm as extracted from the $\Delta\omega$ from the relaxation dispersion curves (Table 2.9.4.2).
Residues Leu213, Gly217 both show slower then bulk exchange rates, $k_{ex}^{-1} \sim 3$ ms and with a p_a of ~ 0.85. These residues are located in β-sheets, β9, β10, which are located at the "back" of the interaction site with MalE. Motion seen in this region in the MD simulations (section 2.8) could suggest that this region undergoes

conformational changes on to the binding surface of MalE. Multiple binding modes of interaction for MalF-P2 to MalE have been suggested in earlier work by Schneider and co-workers[48].
Also, Arg180 shows an exchange rate of 4.53 ms with a p_a of 0.80. R180 is located direct on the kink of α-helix α2 as seen from the crystal structure of $MalFGK_2$[45]. In solution, this α-helix takes the form of an extended α-helix with no kink. This α-helix can thus be seen undergoing a conformational change introducing this kink in the middle of the α-helix due to the interaction with MalE.

3 Discussion

In this section the experimental results of MalF-P2 and MalF-P2 in complex with MalE are discussed.

3.1 Structure and function of soluble MalF-P2

The exceptional length of the second periplasmic loop in the transmembrane protein MalF of the maltose transporter is a particular property of *enterobacteria*[134] and deletion of this region impairs maltose transport. So far it has not been understood why *enterobacteria* require the extension of this loop. MalF-P2 adopts an Ig-like domain structure and can be considered as a receptor domain for MalE. Given the fact that MalF-P2 has no stabilizing contacts to any protein component other than MalE in the X-ray structure of the maltose transporter, we can assume that the isolated MalF-P2 is a good model to study the interactions between the maltose transporter and MalE.

MalF-P2 was purified as a soluble polypeptide (residues MalF91-274) and initial CD measurements gave indications of a well-folded protein with a mixture of α-helix and β-sheet secondary structure, see section 2.2.1. AUC experiments could confirm that MalF-P2 is in a monomeric state at the concentration range used for all experiments in this study (0.1 – 2 mM). Uniform ^{15}N labelled polypeptide yielded a well-resolved ^{1}H, ^{15}N-HSQC spectrum with an almost complete set of resonances for the polypeptide. These early experiments all showed that an extensive NMR study is feasible for this system. Chemical shift assignments could be achieved by uniform labelling of MalF-P2 with ^{2}H, ^{13}C, ^{15}N. Perdeuteration of the polypeptide was a requisite for more complicated 3D experiments to work. The more sensitive 3D experiments, HNCA, HNCO could however be recorded without perdeuteration. Complete assignment could be achieved for the backbone of MalF-P2 with the exception of two segments, the first eight N-terminal residues, and at the C-terminal end, residues 258 – 268). Secondary structure predictions for MalF-P2 in solution, from Cα/Cβ, H^N and N chemical shifts, showed that the two domains of MalF-P2 (domain 1: residues 91-113; 209-259, domain 2: residues 117-207) seen in the $MalFGK_2$-E crystal structure correspond well to the structured regions observed for the isolated MalF-P2. Four β-strands seen in the X-ray structure have not been confirmed from the NMR data. The three short β-strands (β1, β5, β8) are discarded as secondary structures from the NMR data (β-strands need three subsequent residues to be considered a β-strand), and β11 cannot be fully confirmed due to missing Cβ chemical shifts at both ends of the β-strand. Further, the α-helix at the C-termini, α3, showed no α-helical propensity in solution. This α-helix appears to be an amphipatic helix in the X-ray structure, where the aromatic rings of F261, F264 and W268 are presumably embedded into the lipid membrane. This α-helix can then possess intricacy to fold up fully in solution without the lipid membrane or the other parts of the TM regions contributing from other parts of the full transporter. Moreover, analysis of measured RDCs shows that the H^N-N and Hα-Cα bond vectors of the two domains of MalF-P2 yield a high coefficient individually with the crystal structure,

with R-values for domain 1: 0.78 and 0.81 and domain 2: 0.83 and 0.87 for H^N-N and Hα-Cα, respectively. The R-values for the RDCs fitted to the full structure would be an average of the two domains if the two domains were oriented as in the crystal structure. This is however not the case, instead the R-values for the H^N-N and Hα-Cα bond vectors for the full structure decrease to 0.68 and 0.58 for H^N-N and Hα-Cα, respectively. The two domains are therefore not oriented in the same manner in solution as in the crystal structure. This is not unexpected. NMR relaxation experiments of the N^H atoms show that the 20 kDa MalF-P2 protein has a correlation time, τ_c = 8.4 ns, which corresponds to a globular 15 kDa protein in solution. One explanation to this would be that the two domains could tumble independently from each other and hence show increased dynamics towards a globular protein. The effect of this higher than expected tumbling rate is that the linewidth in the NMR spectrum are sharper than expected (the linewidth depends directly on the T_1 and T_2 relaxation rates. This is very advantageous and increases the resolution in crowded spectral regions.

As of now, no high-resolution 3D structure is available for MalF-P2, as the crystal structure has an rmsd of 3 - 5 Å in this region[45] and is missing in the resting state structure[87]. A NMR structure could not be obtained so far. Distance restraints for NMR structural constraints were very scarce (~3500 ambiguous and ~1500 unambiguous NOEs). Sidechain assignments could not be readily assigned due to spectral overlap, where sidechains adopted random coil values. This increases the ambiguity of the NOE data greatly. Typical values for a high resolution NMR structure are in the range of 30 NOE distance restraints per residue. This has the effect that principally all NOEs in the experimental data sets needs to be assigned for MalF-P2. The low number of distance restraints is most likely due to highly flexible regions (giving overlap of the sidechain chemical shifts) and missing long-range contacts within the protein (low number of distance restraints gives no anchor points for the two domains in the structural annealing protocol). Whether this is a problem intrinsic to the construct under study or whether these difficulties can be circumvented by different labelling schemes is not yet known.

Merino *et al.* predicted that both conformations of MalE (maltose-free and maltose-bound) can interact with the maltose transporter[135]. The authors showed that maltose-bound MalE competes with maltose-free MalE in transport assays, and by fitting their data to theoretical equations, they conclud that maltose-free and maltose-bound MalE interact with the transporter with similar affinity. Hence, both forms of the binding protein have access to the membrane components. Recent cross-linking experiments demonstrating close proximity of MalE(G13) to the first periplasmic loop of MalG(P78) in the presence and absence of maltose are consistent with this notion[47]. Also, the interaction between MalF-P2 and maltose-free and maltose-bound MalE could be corroborated to the X-ray structure by ^{1}H, ^{15}N-HSQC experiments with no apparent change in affinity for the maltose concentration in solution. The affinity between MalF-P2 and MalE could be determined to be on the order of 10 - 20 μM by ITC measurements, independent of the maltose concentration. The crystal structures implied that binding of MalE to the transporter might be independent of maltose after

superpositioning of the crystal structures of liganded and un-liganded MalE (pdb codes; 1mep, 1mpd, respectively), as the interaction surface between MalF-P2 and MalE is unchanged. The N-terminal domain can be seen having the same surface in the ribbon diagram of maltose-free and maltose-bound MalE. The rather low affinity could explain why co-elution assays with the two proteins were unsuccessful (experiments performed by Mathias Grote, Humboldt, Berlin). ITC binding data reflects an artificial situation, encompassing only minor parts of a larger system *in vitro*, and is not necessarily an adequate expression of the affinity of MalE to the full complex. However, Davidson and co-workers found that there was no change in affinity of MalE to $MalFGK_2$ in its maltose-free or maltose-bound form, which is in agreement with NMR and ITC results[136]. Also, the ITC determined affinity fits well to the results of Austermuhle *et al*.[44] and Daus *et al*.[47]. Interacting amino acids could be thereby be determined using ^{1}H, ^{15}N-HSQC titration experiments and identifying chemical shift perturbations upon titration of ^{15}N labelled MalF-P2 to MalE. The identified amino acids correspond to observed interactions in the X-ray structure. This indicates that this mode of binding reflects the physiological situation accurately. Cross-linking experiments for the full-length transporter with MalE could also confirm the NMR interaction data (performed by Mathias Grote, Humboldt, Berlin). Moreover, it seems as if the interaction of MalF-P2 to MalE is in accordance with the catalytic intermediate of the crystallized transporter and for the soluble P2 construct, as well as being independent of maltose.
All amino acids in the P2 domain of MalF that could be cross-linked to MalE in the $MalFGK_2$ complex[47], map to the interaction patches identified by NMR. In this study cross-linking experiments were conducted to determine distinct interacting sites, on the full-length transporter. Whereas this makes perfectly sense for the mutation on MalF177, which reacts in the apo-state and in several other conformers of the transporter, this is somewhat surprising for mutations on positions 205 and 252, which cross-link only in the ATP-bound and the vanadate trapped states or the vanadate-trapped state alone, respectively. In line with the tryptic digestion studies[48], this could point to the fact that positions in the C-terminal half of MalF-P2 are less exposed in the ground state of the full transporter than in its catalytic intermediate. This shielded or refolded conformation may not exist in the soluble P2 domain, leading to an exposure of larger interaction patches with MalE. However, this hypothesis has to be taken with precaution, as the cross-linking experiments are not identical for all of the residues considered: Whereas site-specific homobifunctional linkers were employed for the MalF177 position, heterobifunctional photo cross linkers were used for the others. Thus, theoretically one could imagine that MalF177 is even permanently in contact with other epitopes in MalE that could not be detected.

3.1.1 The interaction between MalF-P2 and MalE

Interaction between MalF-P2 and MalE could be observed first with CD difference spectra and were later characterized in more detail by ITC and NMR measurements. Even if ITC cannot give residue specific data, it can provide valuable information about the strength of the interaction. NMR is on the other hand one of the few methods capable of providing interaction data with atomic resolution.

From the ITC data, a dissociation constant, K_D = 10 – 20 µM, could be determined, which is in the range of favourable conditions for NMR studies. Weaker interactions (mM range) can generally not be detected due to the fast exchange of the ligand. Due to the quick dissociation the chemical shifts are not strongly affected and are difficult to detect. Strong interactions (nm range) on the other hand exchange slowly and cause additional resonances, belonging to the bound state. These additional resonances must be assigned to interpret the data. Also, lineshape analysis of slow exchanging systems is tedious and difficult to interpret. Upon a 10:1 molar ratio of MalF-P2 to MalE small perturbations of the chemical shifts can already be observed. At a 1:1 equimolar ratio, most of the resonances show a clear decrease in intensity and a specific subset displays significant chemical shift changes. This indicates that these amino acids are involved in MalE binding. Most notable is the signal decrease upon titration of MalE. This occurs over the major part of the primary sequence. The signal intensity at a 1:1 molar ratio of MalF-P2 and MalE is decreased 5-fold. Only the C-terminal amino acids remain largely unaffected, indicating that these residues do not take part in the interaction. Secondly, a 2-fold increase in linewidth can be observed for those residues, which are affected by an intensity decrease, indicative of that the relaxation rates have changed. These two effects are in accordance with a larger complex being formed with MalF-P2.

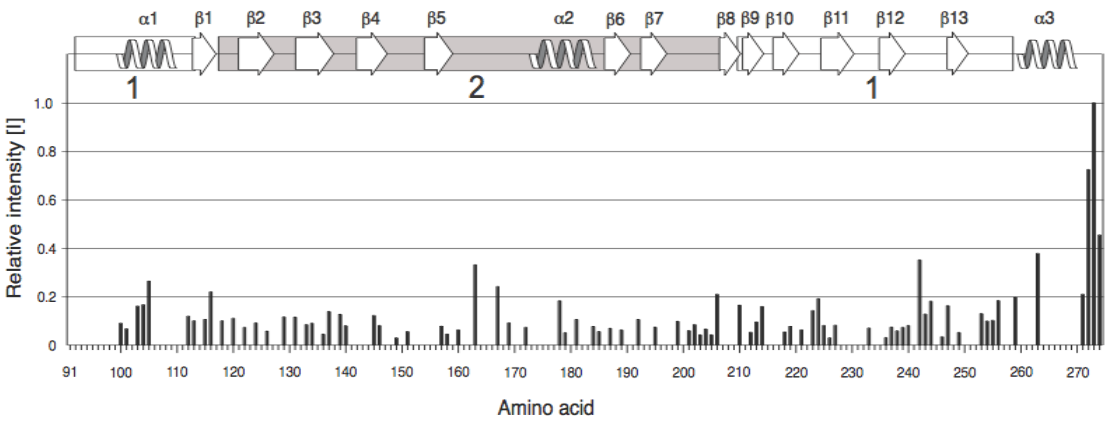

Figure 3.2.1. Relative signal intensity of free MalF-P2 and MalF-P2 in complex with MalE (1:1 molar ratio).

The intensities of bound MalF-P2 are normalised to K273 of free MalF-P2. This residue shows similar linewidth in the free and bound MalF-P2 spectra. Secondary structure elements as determined from the X-ray structure are plotted on the top together with the domain boarders for domain 1 and 2 of MalF-P2. Signal intensities are roughly 5-fold lower over the full primary sequence after addition of equimolar amounts of MalE.

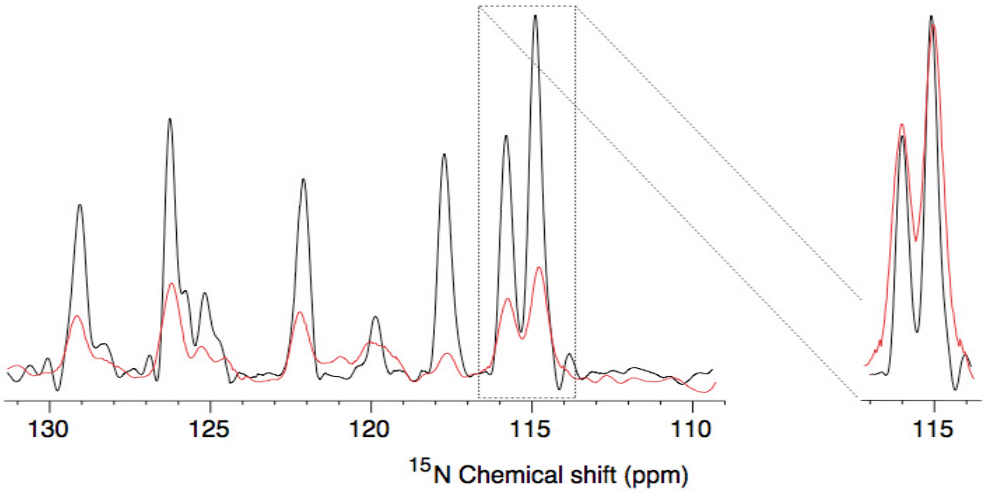

Figure 3.2.2. ^{15}N signal intensity and linewidth comparison upon binding of MalE to MalF-P2.
^{15}N chemical shift comparison of MalF-P2 (black line) and the MalF-P2 – MalE complex (red line). Upon equimolar titration of MalE to MalF-P2 the signal intensity decrease 4-fold. Also, the linewidth is increased by 2-fold. Zoomed in region have intensities scaled to better highlight the linewidth increase (36 Hz MalF-P2 and 64 Hz MalF-P2 – MalE complex).

Using the experimentally determined affinity for MalF-P2 and MalE, the concentration of the complex [MalF-P2: MalE] can be calculated from the concentration of MalF-P2 and MalE in solution. From the rate constant, considering a second-order reaction of two first-order reactants[131], the concentration of the MalF-P2 - MalE complex can be evaluated by:

$$[ES]_i = \frac{1}{2}\left([E]_i^0 + [S]_i^0 + \frac{1}{K} - \sqrt{\left([E]_i^0 + [S]_i^0 + \frac{1}{K} \right)^2 - 4[E]_i^0 [S]_i^0} \right) \qquad [3.2.1]$$

At a 1:1 molar ratio, with a 1 to 1 stoichiometry for the complex formation and a 0.7 mM protein concentration of MalF-P2 and MalE, the concentration of [MalF-P2: MalE] ([ES]) can be determined to be 0.69 mM (99% of MalF-P2 molecules are involved in the complex formation). This is in agreement with the decrease in signal intensity of the NMR spectra.

The effect of the exchange process on the NMR spectrum depends on the difference in the chemical shift frequencies of the two sites. On a timescale of the chemical shift difference of the free and complex state, they compare to the exchange rate, i.e. the chemical shift of a specific residue in its bound and free state. If the exchange rate constant has roughly the same magnitude as the chemical shift difference frequency, the system is said to be in the intermediate exchange regime[137]. This is the regime in which the NMR lineshape are particular sensitive to the chemical exchange process. Considering a symmetric two-site exchange system, with extremely rapid transition between states A and B, so that no intermediate states are involved. The rate constants for the reactions A $\leftrightarrow$ B are considered to be equal (if populations are equal as well!),

and are denoted κ. This means that each molecule has a probability κτ of making a transition to the other state within a short interval τ. The rate constant κ is also known as the *transition probability per unit frame*[137]. In practice this means that unpredictable jumps between states causes the precession frequency to jump and enhanced dephasing of the transverse magnetization occur. This loss of coordination leads to enhanced decay of the total transverse magnetization, which directly affects the lineshapes. These motions can be extracted from the NMR data and be used to acquire the exchange rate between the two states or as in this case the dynamics of the interaction between MalF-P2 and MalE. Different ways have been proposed to extract the dynamics from the NMR data in recent years; either by lineshape analysis[138] or by slow intermediate exchange type experiments[139].

Analysis of the interaction sites from the NMR data with the X-ray data are in high agreement and shows interactions of both domains of MalF-P2 to MalE.
Differences between NMR and X-ray data can be seen for the β-sheets; β1, β2, β3, β4, β6 and β12. Only β-sheet β1 is in close proximity in the crystal structure explaining its change in chemical shift with its interaction with MalE. β-sheets β2, β3, β4, β6 and β12 are all at >5 Å distance from MalE in the crystal structure. Their change in chemical shift could be explained through the hydrogen bonding in the β-sheets. Upon interaction of an amino acid with MalE the perturbation of its secondary structure will propagate through the hydrogen bonds in the β-sheet and induce chemical shift changes at large distances in space. Chemical shift differences can be seen over the β-sheets at hydrogen bonding sites of the backbone. Moreover, the large loop following β7 in the crystal structure is in close vicinity of MalE. The loop contains two methionines that are sensitive to oxidation and could possibly interfere with MalE binding. Non-confirmed data also give indication that the interaction between MalF-P2 and MalE weakens, after specific oxidation of the Methionines. Additional NMR and ITC data need to be collected to confirm this.

Overall MalF-P2 has two interaction surfaces that mainly involve the two α-helices in their respective domains of MalF-P2. Whether both interaction surfaces are needed to accommodate the MalE binding are not yet understood. Cross-linking and tryptic digestion studies performed by Daus *et al*.[47,48] of MalE to differently trapped states of $MalFGK_2$, indicates multiple conformations of MalF-P2 to MalE in the catalytic cycle. To further address this possibility, mutation studies of MalF-P2 encompassing binding studies of both domains separately needs to be conducted.

3.1.2 Structural changes of MalF-P2 upon binding to MalE

From the crystal structures Oldham *et al*.[45] and Khare *et al*.[87] of the $MalFGK_2$ complex in its resting- and catalytic-intermediate state, much can be learned about the mechanical movements of the transporter to accommodate maltose transport through the periplasmic membrane. The two crystal structures unfortunately lack data of the dynamic regions of the periplasmic interface. The second periplasmic loop P2 of MalF can only be seen in the catalytic intermediate structure, where it is tightly bound

to MalE. On the other hand it is also in close contact to MalK subunits in the symmetry related unit cell. In the resting state, only the N-terminal and C-terminal α-helices α1 and α3 yield electron densities. A clear understanding of MalF-P2 relevance in the transport cycle is therefore not clear.

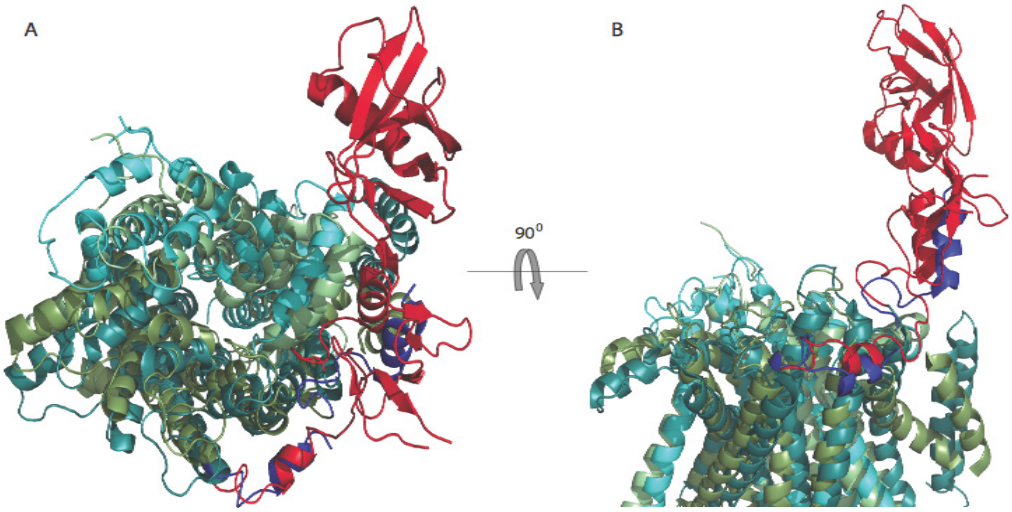

Figure 3.3.1. Structural overlay of the maltose transporter in its resting and an catalytic intermediate state.
Structural overlay of the maltose transporter in its resting-state[87] (pdb code 3fh6, cyan ribbon) and in an intermediate catalytic state[45] (pdb code 2r6g, green ribbon). MalF-P2 is depicted in blue in the resting state and red in the catalytic intermediate state. The figure was made with PyMol[55].

From the crystal structures of $MalFGK_2$, the two domains of MalF-P2 have freedom to undergo different subsets of conformations. This was also speculated from the NMR RDC data, where a grip-claw motion was postulated upon binding of MalE to MalF-P2[107]. The full transport cycle cannot, however, be probed with the MalE – MalF-P2 system in solution. Upon modelling of the structures of liganded and unliganded MalE (pdb 1mpd and 1peb, respectively), aligned to the MalF-P2 – MalE interaction surface, severe steric clashes (Figure 3.3.2) occur. The liganded MalE must hence be able to bind to the $MalFGK_2$ in a yet undiscovered way.

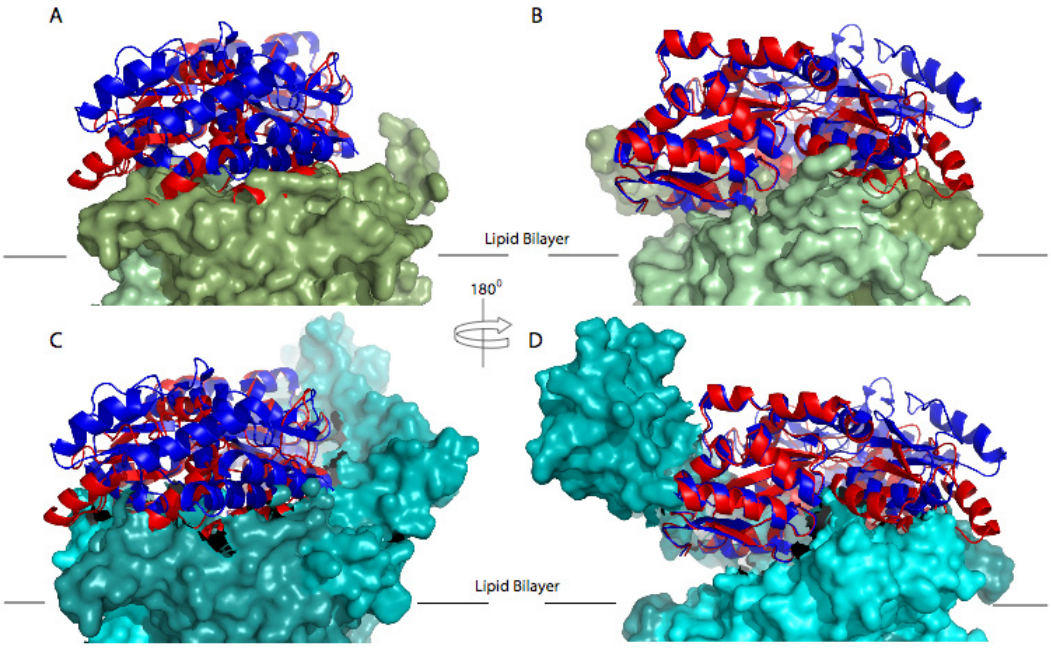

Figure 3.3.2. Liganded MalE modelled into the resting and catalytic intermediate state of MalFGK2. MalE modelled into the crystal structures of the resting state[87] (panels A and B, pdb code 3fh6, green surface) and the catalytic intermediate state[45] (panels B and C, pdb code 2r6g, cyan surface) of the maltose transporter. Panel A and B depicts the resting state structure with unliganded MalE (blue ribbon) and the liganded MalE (red ribbon; pdb code 1mpd) modelled in. Panel C and D show the catalytic intermediate state (MalF, MalG in cyan surface, MalE blue ribbon) with liganded MalE modelled in (red ribbon; pdb code 1mpd). For all panels, liganded MalE is aligned to the N-terminal part of MalE in the intermediate complex. Large sterical clashes are observed for liganded MalE and MalG both in the intermediate and the resting state of the periplasmic interface of $MalFGK_2$.

Different binding modes are also observed in the fitting of the RDC data in the bound and unbound state of MalF-P2. H^N-N RDC data of domain 1 for the bound and unbound MalF-P2 have, as expected, a higher correlation to the crystal structure upon binding of MalE, R = 0.78 for free MalF-P2 and R = 0.90 for MalF-P2 in comlex with MalE. For domain 2 this is however not the case, the correlation of the H^N-N data fits worse upon complex formation of MalF-P2 and MalE, R = 0.87 for free MalF-P2 and R = 0.52 of MalF-P2 in complex with MalE. The reasons for this behaviour could be due to different factors, i) crystal contacts might induce local structure perturbations, which in turn effect the local conformation of domain 2. ii) interaction of domain 2 to the MalK subunits in the symmetry related unit cell, might be preferred over the interaction with MalE. Causing MalF-P2 to take a structure more related to the unbound state. iii) The bound structure of domain 2 of MalF-P2 cannot be fully accessed by the crystal structure since it cannot accommodate the true binding interaction within the crystal. However, the H^N-N RDC data of the bound state confirm that the overall orientation of the two domains is indeed the same in solution

and in the crystal structure. Unfortunately structure refinement calculations of MalF-P2 with H^N-N RDC data of MalF-P2 in complex with MalE were so far not successful. To improve structural refinement multiple sets of RDC are needed in the bound state. By using multiple alignment media to lower the degeneracy of the RDC data, the problem of acquiring multiple RDC data sets could be circumvented.

3.1.3 Binding kinetics between MalF-P2 and MalE

For a more complete dynamic view of MalF-P2 in solution and in complex with MalE, a CPMG based approach by NMR was performed. CPMG experiments are sensitive on the μs – ms timescale, which provides information about the conformational changes in ligand complexes and kinetic processes[121-123].
CPMG experiments in solution show that MalF-P2 undergoes conformational changes on the ms time range. A second population, p_b, of 1% of total molecules in solution, exchanges with a bulk time of 0.738 ms. This implies that the two domains of MalF-P2 can reorient themselves to each other in solution. From the determined scaling factor α, this exchange occurs on the slow exchange regime where two populations exist and the low population cannot be detected.
In the complex of MalF-P2 and MalE the exchange of MalF-P2 from its bound to un-bound state can be determined from the bulk exchange rate from the dispersion curves. The lifetime of the complex is on a time scale of 0.538 ms.
The slow dynamics of the MalF-P2 – MalE complex might explain the problems in fitting RDC data of MalF-P2 bound to MalE. The experimental RDCs represent the average H^N-N bond vectors of the studied molecule. When determined H^N-N bond vectors of the average structure are compared to static structural data (i.e. X-ray structures), a large discrepancy can be obtained. In the case of the $MalFGK_2$-E crystal structure[45], domain 2 of MalF-P2 is in close contact of two different MalK subunits in the symmetry related unit cells. These contacts may in addition cause a divergence in the H^N-N bond vectors to those measured in solution. Indeed, the RDCs measured of the MalF-P2 – MalE complex differ for domain 2 but not for domain 1, which yields a high correlation. To obtain a high-resolution structure using RDC refinements, a large set of bond vectors need to be determined or multiple alignment media to reduce the degeneracy. It was not possible to use necessary 3D experiments to obtain multiple sets of RDC data for the MalF-P2 – MalE complex due to its size. This emphasizes the importance of understanding the dynamics of the studied system when comparing static with dynamic data.
NMR data show that α-helix α2 of MalF-P2 is a well-folded elongated α-helix in solution. The X-ray structure on the other hand shows α2 as an elongated α-helix with a kink in the middle. It was not confirmed from either NMR or X-ray data if this was an artifact from one or the other studies. The relaxation dispersion profile of Arg180, located directly on the kink, shows a large deviation from the bulk exchange rates and population distribution. The conformational change might be related to the interaction between MalF-P2 and MalE, which introduces this kink. MD simulations of the complex show that MalF-P2 undergoes large-scale dynamics on the interface

surface, always retaining the kink of the α-helix. This structural feature is clearly needed to accommodate the association between α-helix α2 and MalE.
Furthermore, residues Leu213 and Gly217 located in β-sheets β9 and β10 at the "back" of the interaction surface in domain 1 of MalF-P2, shows a large deviation from bulk exchange rates and population distribution (section 2.5.4.4). These β-sheets are not involved in the interaction to MalE, but nevertheless show dependence. The motions seen in relaxation dispersion curves for β-sheets β9 and β10 are in agreement with the X-ray structure, with higher B-factors. The μs – ms timescale is highly unfavorable for X-ray diffraction yielding very poor electron density data. The crystal structure is either missing or is at a resolution of >4 Å (TMD regions having a average resolution of 3.0 Å) for this region[45]. The cause and function of this dynamic behavior is currently not understood.

3.2 General discussion

ATP-binding cassette (ABC) transporters constitute a ubiquitous superfamily of integral membrane proteins that are responsible for the ATP-translocation of many substrates across membranes. ABC transporters are involved in tumour resistance, cystic fibrosis, bacterial multidrug resistance, and a range of other inherited human disease. This makes them viable drug targets for a range of diseases.
The highly conserved ABC domains of ABC transporters provide the nucleotide-dependent engine that drives transport. By contrast, the transmembrane domains that create the translocation pathway are more variable[39]. As has been suggested by several groups, binding of substrate initiates the transport cycle. The "power stroke", that is, ATP binding that induces NBD dimerization and formation of the ATP sandwich, drive the conformational changes in the TMDs[64]. Intra- and extracellular loops connecting the TM segments of ABC transporters are of varying lengths and possess functional significance for contacts to the NBDs or extracellular interaction partners. The "coupling helices" are architecturally conserved elements forming the NBD-TMD interface[52]. In the crystal structures, they are located in grooves on the NBDs' surfaces. Nucleotide-induced conformational changes of the NBDs are thought to be transmitted via non-covalent interactions from the Q-loop region of the NBDs to these short helices.
Periplasmic binding proteins (BPs) are essential components of ABC-importers. Substrates bound by BPs are highly diverse, ranging from inorganic or organic ions and sugars to peptides or vitamins. These proteins are needed at very high substrate concentrations, and may also play an important functional role in the catalytic cycle of the transporter[140].
Given the vast prevalence of ABC transporters throughout life, surprisingly few have been characterized, either biochemically, structurally or mechanistically. Therefore a broad generalization about their function is extrapolated from just a few observations.
The maltose importer $MalFGK_2$-E of *E.coli* is probably the best characterized ABC transporter to date[37,45,87,141,142]. The crystal structures of the full transporter in the resting state[87] and a catalytic intermediate state[45], or a solution state paramagnetic

NMR study of maltose binding protein[143] show that $MalFGK_2$-E can be considered as a model system for research on ABC transporters.
In this work the unusually large periplasmic loop P2 of MalF has been under careful investigation. The periplasmic loops are crucial for transducing hydrolysis and binding events. The large second periplasmic loop of MalF (MalF-P2) is unique for *enterobacteria*. Schneider and co-workers demonstrated that MalF-P2 undergoes an ATP-dependent movement, and such that MalF-P2 reorients itself with respect to the other subunits of the ABC transporter in the course of the transport cycle[48]. In fact, MalF-P2 seems to act as a receptor, which recruits MalE, and thus maltose, to the pore of the membrane protein.
The question how much MalF-P2 contributes to substrate recognition is still unresolved. Earlier genetic and biochemical investigations show that both MalF and MalG are involved in binding to the substrate-binding protein MalE[47,144]. It could be demonstrated that the N-terminal lobe of MalE interacts with MalG, preferentially to the P3-loop, whereas the C-terminal part of MalE is close in space to MalF (MalF-P2)[45]. This study shows that MalF-P2 interacts with MalE independent of maltose and undergoes large conformational changes in the interaction. The binding interface of MalE to MalF-P2 is analogous independent of maltose. In contrast, MalF-P2 is highly dynamic in its interaction to MalE and can sample multiple conformations on the binding interface. This might be a necessity to maintain the interaction during the catalytic transport of the full length maltose transporter.
In conclusion, this study demonstrates that the P2-loop of MalF interacts with MalE in all states of the transporter during the catalytic cycle, yet in a differentiated manner.

It is tempting to speculate that MalF-P2 could act as the receptor for substrate recognition through MalE and during the interaction induce MalE to release maltose into the transporter pore. Earlier mutational studies of MalE show that the affinity of MalE and maltose can be introduced by minor mutations far away from the substrate binding site[145,146]. These mutations are located on the outward facing hinge of MalE. These “hot spots” are in contact with MalF-P2 upon binding. As discussed earlier in section 3.1.3 the fully closed MalE cannot be modelled into either the resting state or catalytic intermediate state of the $MalFGK_2$ transporter. If MalF-P2 on the other hand can induce opening of MalE, the fully assembled system of $MalFGK_2$-E described by Oldham *et al.*[45] could be accomplished. This is on the other hand very speculative and needs to be further pursued.
However, whether these above mentioned interactions alone are sufficient to trigger the proposed events or whether they have to be considered together with other molecular interactions to induce the transition towards another functional state, remains to be elucidated.

4. Materials and Methods

This chapter describes the methods and experimental setup used for MalF-P2 and the MalF-P2 – MalE complex.

4.1 Chemicals and supplementary equipment

If not indicated otherwise, all chemicals and NMR material were purchased from Amersham Biosciences (Braunschweig, D), Biomol (Hamburg, D), Bruker Biosciences (Karlsruhe, D), Profos AG (Regensburg, D), Deutero (Kastellaun, D) Fluka (Neu-Ulm, D), Merck (Darmstadt, D), Roth (Karlsruhe, D), Sigma (Deisenhofen, D).

4.2 Bacterial strains and plasmids

MalF-P2 was overproduced in *E.coli* BL21 DE3 <pLysS> harbouring the plasmid pMG15.
MalE was overproduced in *E.coli* BL21 DE3 <pLysS> harbouring the plasmid pLic-His_6.
Both expression systems contain carbenicillin antibiotic resistance and *E.coli* BL21 DE3 <pLysS> contains chloramphenicol antibiotic resistance.

4.3 Media

LB (Luria-Bertani) was prepared as described by Sambrook *et al.*[147]. Isotope enriched protein was achieved by growth in M9 minimal media supplemented with 0.05% $^{15}NH_4Cl$ and/or 0.2% ^{13}C-glucose (Cambridge Isotope Laboratories, Andover, USA), 1mM $MgSO_4$, 0.5 mM $CaCl_2$, biotin (1mg/ml), thiamine (1mg/ml) and trace elements. For full deuteration of the protein 0.2% $^{2}H^{13}$C-glucose (Cambridge Isotope Laboratories, Andover, USA) and 99.9% D_20 (Euroisotop, Gif-sur-Yvette, F) were supplemented into the M9 minimal media.
Media were added with chloramphenicol (20 μg/ml) and/or carbenicillin (100 μg/ml).

4.4 Cell growth and storage

Cell cultures were prepared from fresh transformation of the plasmids into the competent cells of its *E.coli* strain. Seven colonies were inoculated from an agar plate and were screened for expression. Pre-cultures were then inoculated from the highest expressing colony. Pre-cultures were grown under aerobic conditions at 37 ^{0}C. Cell growth was monitored photometrically at a wavelength of 600 nm (S800, WPA, Cambridge, UK).
After cell lysis the lysate was washed with 20mM phosphate buffer with 100 mM NaCl, pH 7.4, with appropriate amounts of protease inhibitor cocktail tablets (Roche, Mannheim, D, 1x50ml buffer volume). The cell lysate was then stored at -80 ^{0}C until further purification.

4.5 Molecular Biology

4.5.1 Plasmid preparation and amplification

Plasmids were prepared with the "NucleoSpin Plasmid"-kit according to the manufacturer's instructions (Macherey-Nagel, Düren, D).
Plasmid amplification were made in competent GIGA cells and extracted with Wizard^{+} Midiprep (Promega Madison, Wi, USA).

4.5.2 DNA electrophoresis

Electrophoresis of DNA and visualization of bands was essentially performed as described by Sambrook *et al.*[147]. "1 kb plus DNAladder" (Invitrogen, Karlsruhe, D) was used as a marker. Electrophoresis was performed by Mathias grote (Humboldt, Berlin).

4.5.3 Polymerase Chain Reactions (PCR)

Amplification of DNA-fragments was performed by PCR using site-specific oligonucleotides (Invitrogen, Karlsruhe, D) and plasmid DNAas template. *Pfu*ReproFast-DNA-Polymerase (Genaxxon, Munich, D) was used and PCR-assays were pipetted according to the instructions. After subcloning of the fragments into plasmids, the DNAwas routinely sequenced (MWG Biotech, Martinsried, D).
For the cloning of the P2-domain of *E.coli malF*, the following oligonucleotides were used:

malF_pploop_fw
5'-GCTCTACATATGAACTACAGCAGCACTAAC-3'

malF_pploop_rv
5'-CCATGTGGATCCTTATTTCTGAATGCCTTCGTCGGTAAAGACG-3'

These primers were designed to comprise NdeI/BamH1 restriction sites and a 6 base overhang. The following PCR cycling was used (30 cycles):

95° 2 min
95° 30 min
58° 1 min
72° 35 min

No extra $MgCl_2$ was added.
PCR reactions was performed by Mathias Grote (Humboldt, Berlin)

4.5.4 Competent cells and transformation

Transformation of BL21 DE3 <pLysS> competent cells (Stratagene, La Jolla, Ca, USA) was done according to manufacturer's instructions. After transformation, 1ml of LB was added and cells were grown aerobically for 1 h at 37 ^{0}C. Cells were then plated on antibiotic containing agar plates over night at 37 ^{0}C.
Electrocompetent cells were prepared according to the protocol of the electroporator "Easyject" (Eurogentec, Cologne, D). 50 µl electrocompetent JM109 cells were mixed with 0.5-1 µl of plasmid DNA or 5 µl ligation assay, incubated for 1 min on ice and transferred to an electroporation cuvette (Peqlab, Erlangen, D). After electroporation, 1 ml of cooled SOC-medium was added and cells were grown aerobically for 1 h at 37 °C. Cells were then plated on selective agar and grown overnight at 37 °C.

4.6 Analytical procedures

Analytical methods used for MalF-P2 and MalE is presented in the next five sections.

4.6.1 Determination of concentrations

MalE and MalF-P2 were quantified photometrically

$$\varepsilon 280(\text{MalE}) = 68750\ M^{-1}cm^{-1}$$
$$\varepsilon 280(\text{MalF-P2}) = 33920\ M^{-1}cm^{-1}$$

Extinction coefficients for MalF-P2 and MalE were determined with ProtParam (www.expasy.org/tools/prot param.html)[148] and verified experimentally.

4.6.2 Circular dichroism (CD-) spectroscopy

Secondary structure of soluble MalF-P2 protein was assessed at a J-720 CD-spectrometer (Jasco, Groß-Umstadt, D). Protein concentration was 0.1 mg/ml (in 10 mM phosphate buffer pH 7.4). Wavelength scans from 260 -185 nm were accumulated 60x at RT, using a bandwidth of 4 nm and with a cuvette diameter of 1 mm.

4.6.3 SDS-PAGE

Polyacrylamide gels were prepared with stacking gels (pH 6.7), containing 4.9 % acrylamide and 0.13% bis-acrylamide, resolving gels (pH 8.4) 15% or 20% acrylamide and 0.4% or 0.55% bis-acrylamide, respectively.
Samples were mixed with sample buffer (4% w/v SDS, 25% stacking gel buffer, 30% v/v glycerol, 10% v/v β-mercaptoethanol, 0.002% w/v bromophenol blue). For determination of molecular weights the protein standard GlowMarker Kit (GE Healthcare, D) was used.

Protein bands were fixed (50% v/v ethanol, 10% acetic acid, 30 min) and stained with Coomassie (0.02% w/v Coomassie Brilliant Blue, 10% acetic acid, 30 min). Excess staining was removed by incubation with 10% acetic acid.

4.6.4 Analytical Ultra Centrifugation (AUC)

Sedimentation analyses of MalF-P2 were performed on a Beckman Optima XL-I analytical ultracentrifuge. All experiments were performed in PBS (pH 7.4). Sedimentation velocity experiments over a range of starting protein concentrations of 100–500 μM were performed by using interference and absorption optics (continuous mode; 0.003 cm step size) at appropriate wavelengths for detection of sedimentation profiles. The samples were analyzed at 20 °C at 40,000–50,000 rpm in a two-sector epon/charcoal centerpiece in the AN-60-TI rotor. Experimental data were analyzed for protein purity, aggregates, and oligomerization status by using c(s) distribution analysis as implemented in Sedfit 9.3[149].

4.6.5 Mass spectrometry

MalF-P2 molecular mass verification and sequence analysis were performed by MS/MS mass spectrometry experiments. For molecular mass verification a 98% pure protein solution of MalF-P2 in pure water was used. Sequence analysis for primary sequence verification and methionine oxidation, were performed by Coomassie Blue staining in SDS-PAGE, and followed by in-gel trypsin digestion as described by Czupalla *et al*.[96].
MS experiments were performed on a quadruple orthogonal acceleration time-of-flight mass spectrometer Q-Tof Ultima (Micromass, Manchester, UK) with a nanoelectrospray source (PicoTip spray capillaries; New Objective, Woburn, MA), as delivery system. For protein identification, data were acquired in a data-dependent mode (survey scanning) using one MS scan followed by MS/MS scans of the most abundant peak[150]. The processed MS/MS spectra and MASCOT server version 1.9 (Matrix Science Ltd., London, UK) were used to search in-house against the SwissProt protein database (210906 and 234112 sequences).

4.7 Protein expression, purification and modification

This section contains information for the expression and purification steps for the two proteins used in this study.

4.7.1 Expression and purification of MalF-P2

MalF-P2 was expressed from pMG15 in *E.coli* BL21 DE3 <pLysS>. LB- or minimal media supplemented with carbenicillin and chloramphenicol was inoculated to OD_{600} = 0.05, with fresh overnight culture and incubated at 37 ^{0}C and 250 rpm. Expression were induced at OD_{600} = 0.6-0.8, by the addition of 0.5 mM IPTG, for 4 hrs. Cell growth continued after induction until the late logarithmic phase, OD_{600} = 2-4. Cells were harvested (20 min, 6500 x g) and resuspended in buffer (20mM phosphate pH

7.4, 100 mM NaCl, 1x protease inhibitor tablet) and disrupted by ultrasonification (6x for 1min, level 7 on ice, maximum temperature set to 15 ^{0}C; Branson Sonifier, Schwäbisch Gmund, D). Disrupted cells were centrifuged (45 min, 16000 x g) and supernatant was incubated with equilibrated nickel-matrix (GE Healthcare, Munich, D) at 5 ^{0}C overnight. The nickel column was washed with 10 column volumes of buffer and with 15 columns volume of buffer with 25 mM imidazole. Protein was then eluted with buffer containing 250 mM imidazole. Imidazole was removed by buffer exchange by either dialysis or amicon filters (Millipore, Schwalbach, D) before thrombin digestion. Thrombin digestion of the N-terminal His_6-tag was performed with, 5 units/mg protein, thrombin protease (Amersham Biosciences, Freiburg, D) overnight at 5 ^{0}C in a falcon tube. The protein solution were then dialysed, and as a final purification step, subjected to a Superdex 75 gelfiltration column on a Fast Protein Liquid Chromatography (FPLC, GE Healthcare, Munich, D) to remove residual imidazole, His-tag fragments and any impurities.

4.7.2 Expression and purification of MalE

Wildtype and mutant His_6-MalE were expressed and purified as described by Cabrita *et al.*[95] and Daus *et al.*[94] respectively.

4.7.3 Chemical cross-linking with sulfonate cross-linkers

Crosslinking with homobifunctional sulfonate crosslinkers was performed as described by Daus *et al.*[94]. The following crosslinkers were used:
1,2 ethanediyl-bismethanethiosulfonate (EBS);
1,6 hexandiyl-bismethanethiosulfonate (HBS) and
3,6,9,12,15-pentaoxaheptadecan-1,17-diyl-bis-methanethiosulfonat (PBS). According to Loo & Clarke[151], the approximate spacer lengths are 5.2 Å (EBS), 10.4 Å (HBS) and 24.7 Å (PBS).

4.7.4 Methionine oxidation of MalF-P2

Oxidation of the three methionines of MalF-P2 was achieved by incubation of MalF-P2 with a 1:1, 2:1, 5:1, 10:1, 50:1, 100:1 fold excess of H_2O_2 (4 hrs, room temperature). After incubation samples were frozen and lyophilized and resuspended in buffer (20 mM phosphate pH 7.4, 100 mM NaCl). All samples were checked with mass spectrometry for degradation and size and by ^{1}H solution NMR if fold is maintained.

4.8 Biophysical methods

This section describes the different methods and samples used in the biophysical studies of MalF-P2 and MalE.

4.8.1 NMR

NMR experiments were performed on 0.1-1.0 mM uniformly ^{15}N, $^{13}C/^{15}N$ and $^{2}H/^{13}C/^{15}N$ labeled samples MalF-P2 depending on experimental requirements. Samples for sequence specific assignment of resonances and structure determination were prepared in 20 mM NaH_2PO_4/Na_2HPO_4, pH 7.4, 100 mM NaCl, 0.02% (w/v) NaN3, in 90% H_2O/10% D_2O. Samples for side chain assignments were prepared in the above buffer, lyophilized and then dissolved in 100% D_2O. For NMR based titration experiments with MalE, stock solutions of protein were prepared containing 0.5-2.0 mM of MalE, with same buffer conditions used for MalF-P2.

4.8.1.1 NMR assignments of MalF-P2

NMR data were recorded by using Bruker 600 MHz and 750 MHz spectrometers (Bruker Biospin, Karlsruhe, Germany), both equipped with triple resonance cryogenic probes. For backbone and sidechain assignments of the MalF-P2, standard triple resonance experiments (^{15}N-HSQC, ^{15}N-HSQC-TROSY, HNCA, HNCO, HNCACO, HNCACB, HNCOCACB, HA(CACO)NH, HCCH-TOCSY and HCCH-COSY)[100] were recorded. ^{13}C shifts were adjusted due to the ^{2}H effect according to Venters *et al.*[99]. All recorded experiments are listed in table 4.9.1.1.1. NMR data were processed using XWIN-NMR and/or Topspin (Bruker Biospin, Germany) and analyzed with CCPN[294].
3D ^{15}N-NOESY-HSQC in H_2O[260] and 3D ^{13}C-NOESY-HSQC in D_2O[261] were recorded to further aid in sidechain assignments and secondary structure prediction. All 3D NOESY spectra were recorded at 600 MHz or 900 MHz and the spectra were folded appropriately to increase the resolution of the indirect ^{13}C and ^{15}N dimensions respectively.
Typically 12 transient scans were used for 3D experiments with 128 x 128 x 1024 x points complex points in the $F_1(^{13}C)$, $F_2(^{15}N)$, $F_3(^{1}H)$ dimensions respectively. For NOE experiments a mixing time of 80-120 ms were typically used.
Data was processed with a square sine-bell window function applied in all three dimensions with a SSB range of 2.0 - 3.0, a qfil window function to suppress the water signal and forward/backward linear prediction where applicable to enhance spectral quality. Data was also zero filled to 512 x 512 x 1024 (t_1, t_2, t_3) data points.

4.8.1.2 Secondary structure of MalF-P2

Backbone dihedral angle prediction was performed using ^{15}N, $^{13}C'$, $^{13}C\alpha$, $^{13}C\beta$ chemical shifts of MalF-P2 after correction for ^{2}H isotopic effects[218]. Structural dihedral restraints were generated using the program TALOS[104].
$C\alpha$, $C\beta$ measured chemical shifts were compared to random coil chemical shifts [$\Delta\delta C\alpha = (C\alpha_{,expt} - C\alpha_{,random\ coil})$ and $\Delta\delta C\beta = (C\beta_{,expt} - C\beta_{,random\ coil})$] β-strand secondary structure is characterized by negative $\Delta\delta C\alpha$ and positive $\Delta\delta C\beta$ values for involved residues. α-helices are characterized by positive $\Delta\delta C\alpha$ and negative $\Delta\delta C\beta$ values for the involved residues[102,103]. ^{13}C- and ^{15}N-NOESY-HSQC experiments were

analyzed for secondary structure elements. An α-helical secondary fragment gives strong N^H – CO NOE signals between the amide protons, n; n+3, and also strong NOE signals between N^H-N^H, n; n+1, and can therefore be easily resolved in the ^{15}N-NOESY-HSQC experiments. Antiparallel β-strands give strong Hα- Hα NOE signals between the two strands and can be seen in the ^{13}C-NOESY-HSQC.

4.8.1.3 NMR relaxation experiments with MalF-P2

A variety of experiments used for studying protein dynamics are presented in the following four sections.

4.8.1.3.1 ^{15}N T_1-experiments of MalF-P2

^{1}H-^{5}N T_1 measurements at 600 MHz were recorded with relaxation delays of 12, 52, 102, 152, 202, 302, 402, 602, 902, 2002, 4002 ms. Each ^{15}N-^{1}H HSQC CPMG experiment[102,152] and utilizing a watergate pulsetrain for suppression of the water signal. 256 × 1024 complex points were acquired in the F_1 (^{15}N) and F_2 (^{1}H) dimensions respectively, with spectral widths of 49.5 ppm and 16.6 ppm. 12 transient scans were acquired for each t_1 point, and the acquisition times for the F_1 and F_2 dimensions were 42 ms and 51 ms respectively. Data was processed using sine-bell window function applied in both dimensions. Data was also zero filled to 2048 (t_1) and 4096 (t_2) data points. Relaxation times were obtained from an analysis of the exponential decay of peak intensities using the program CCPN[101].

4.8.1.3.2 ^{15}N T_2-experiments of MalF-P2

^{15}N T_2 measurements at 600 MHz were recorded with relaxation delays of 6, 10, 18, 26, 34, 42, 82, 122, 162, 202, 242 ms. Each ^{15}N-^{1}H HSQC CPMG experiment[102,152] utilizing a watergate pulsetrain for suppression of the water signal. 256 × 1024 complex points were acquired in the F_1 (^{15}N) and F_2 (^{1}H) dimensions respectively, with spectral widths of 49.5 ppm and 16.6 ppm. 12 transient scans were acquired for each t_1 point, and the acquisition times for the F_1 and F_2 dimensions were 37 ms and 150 ms respectively. Data was processed using sine-bell and squared sine-bell window functions applied in the indirect and direct dimensions respectively. Data was also zero filled to 2048 (t_1) and 4096 (t_2) data points. Relaxation times were obtained from an analysis of the exponential decay of peak intensities using the program CCPN[101].

4.8.1.3.3 ^{1}H-^{15}N Heteronuclear NOE measurements with MalF-P2

Heteronuclear ^{1}H -^{15}N NOEs 600 MHz were obtained for MalF-P2 alone, 500 μM sample, and for the MalF-P2 MalE complex, with a 1 mM 1:1 molar ratio, using a gradient sensitivity enhanced ^{1}H -^{15}N NOE experiment[102] recorded with and without ^{1}H saturation as two interleaved data matrices of 256 × 1024 complex points acquired in the F_1 (^{15}N) and F_2 (^{1}H) dimensions respectively, and spectral widths of 49.5 ppm and 16.6 ppm. A relaxation delay of 2 s prior to 3 s of ^{1}H saturation was used for the

NOE spectra, and a 5 s relaxation delay used for the reference spectra. Data was processed using Lorentzian-to-Gaussian filtering functions applied in both dimensions, and zero filling to 2048 (t_1) and 4096 (t_2) data points. The heteronuclear enhancement factor NOE_{600} was calculated from the difference in peak intensities of cross-peaks in the presence (I_{sat}) and absence (I_0) of ^1H saturation according to:

$$NOE_{600} = (I_{sat}-I_0)/I_0, \text{ and}$$

$$SD = \sqrt{[(sd_{sat}\times I_0)^2 + (sd_0\times I_{sat})^2]}/I_0^2,$$

where, the error for each peak is given by the overall standard deviation, SD, of the baseplane noise, and sd_{sat} and sd_0 are the standard deviations of the baseplane noise in the saturated and reference spectra respectively. 32 and 96 transient scans were acquired for MalF-P2 and for the MalF-P2 MalE complex respectively for each t_1 point, and the acquisition times for the F_1 and F_2 dimensions were 51 ms and 42 ms respectively. Data was processed using sine-bell and squared sine-bell window functions applied in the indirect and direct dimensions respectively. Data was also zero filled to 2048 (t_1) and 4096 (t_2) data points. Signal intensities were analyzed and fitted with CCPN[101].

4.8.1.3.4 Carr-Purcell Meiboom-Gill (CPMG)

NMR experiments for slow dynamics were recorded for MalF-P2 and for the MalF-P2/MalE complex. The MalF-P2 were recorded with a uniformly ^{15}N labeled 500 µM sample, and the MalF-P2/MalE complex with a ^{15}N uniformly labeled 500 µM sample with a 1:1 molar ratio.

A ^1H-^5N H(S,M)QC pair were recorded for MalF-P2 and the MalF-P2/MalE complex according to Skrynnikov *et al.*[153], to determine the ^{15}N chemical shift difference due to slow ms dynamics. The ^1H-^5N H(S,M)QC pair were recorded utilizing a watergate pulsetrain for suppression of the water signal instead of the sensitive enhanced echo-antiecho sequence suggested by Skrynnikov *et al.*[153]. 512 × 2048 complex points were acquired in the F_1 (^{15}N) and F_2 (^1H) dimensions respectively, with spectral widths of 49.5 ppm and 16.6 ppm. 16 and 32 transient scans were acquired for each t_1 point for MalF-P2 and for MalF-P2/MalE complex respectively, and the acquisition times for the F_1 and F_2 dimensions were 56 ms and 68 ms respectively. Data was processed using a square sine-bell window function applied in both dimensions. Data was also zero filled to 131072 (t_1) and 2048 (t_2) data points, to obtain a digital resolution of 0.022 Hz and 2.062 HZ respectively. Chemical shift differences in the ^{15}N dimension were determined by lineshape analysis and extracted in CCPN[101].

A ^1H-^5N CT-CPMG pulsetrain were recorded for MalF-P2 and the MalF-P2/MalE complex to determine the slow ms dynamics within the sample, accordingly to Tollinger *et al.*[154]. The ^1H-^5N CT-CPMG were recorded utilizing a watergate pulsetrain for suppression of the water signal instead of the sensitive enhanced echo-antiecho sequence suggested by Tollinger *et al.*[154]. The CPMG train was set to 40 ms

and to 10 ms for MalF-P2 and for MalF-P2/MalE complex respectively, with B_1 fields, υ_{CP}, of 25, 50, 75, 100, 125, 150, 175, 200, 250, 325, 400, 500, 625, 750, 950 Hz for MalF-P2 and 100, 110, 120, 130, 140, 150, 175, 200, 250, 325, 400, 500, 625, 750, 950, 1200 Hz for the MalF-P2/MalE complex. 256 × 2048 complex points were acquired in the F_1 (^{15}N) and F_2 (^{1}H) dimensions respectively, with spectral widths of 49.5 ppm and 16.6 ppm. 24 and 48 transient scans were acquired for each t_1 point for MalF-P2 and for MalF-P2/MalE complex respectively, and the acquisition times for the F_1 and F_2 dimensions were 42 ms and 68 ms respectively. Data was processed using a square sine-bell window function applied in both dimensions. Data was also zero filled to 2048 (t_1) and 4096 (t_2) data points, to obtain a digital resolution of 1.415 Hz and 1.031 HZ respectively. Chemical shift differences in the 15N dimension were determined by lineshape analysis and extracted in CCPN[101].

CPMG data of free MalF-P2 were fitted according to the master equation at two magnetic fields, which is valid at all time-scales provided $p_A >> p_B$. For the MalF-P2 – MalE complex, populations p_A and p_B are calculated from the initial protein concentrations and solving the rate constant equation, considering a second-order reaction of two first-order reactants[131]. Data are then fitted as in the free MalF-P2 case.

4.8.1.4 Titration experiments with MalF-P2

Titration of 1.0 mM MalF-P2 up to 1:8 protein to ligand ratio against MalE was monitored by following the chemical shift change, $\Delta\delta$, in standard ^{15}N-^{1}H HSQC[102] and ^{15}N-^{1}H HSQC-trosy spectra[97], where,

$$\delta = \sqrt{[(\Delta H)^2 + (\Delta N/5)^2]} \qquad [4.9.1.4.1]$$

For each titration point 256 × 1024 complex points were acquired in the F_1 (^{15}N) and F_2 (^{1}H) dimensions respectively, using spectral widths of 49.5 ppm (t_1) and 16.6 ppm (t_2) and 64 transient scans. Comparing the ^{1}H-^{15}N chemical shifts of protein alone to those of the protein, in presence of MalE, the perturbed residues were identified during the titration. Due to lowering of signal intensity upon higher correlation time, T_c, upon binding to MalE, a high number of transient scans needed to be implemented to receive adequate signal to noise. Data was processed using a square sine-bell window function applied in both dimensions. Data was also zero filled to 2048 (t_1) and 4096 (t_2) data points. Chemical shift perturbations were analyzed with CCPN[101].

4.8.1.5 Residual dipolar couplings measurements (RDC)

Residual dipolar couplings (RDCs) ($N_iH_i^N$, $C\alpha_iC'_i$, $H\alpha_iC\alpha_i$) were measured using ^{15}N-HSQC IPAP[155] and *J*-resolved HNCO, HNCA pulse programs[156] for MalF-P2 in its unbound state. ^{15}N-HSQC and ^{15}N-HSQC-Trosy[97] experiments were used to measure the $N_iH_i^N$ couplings in the bound state.
To obtain residual alignment of the protein solutions, dipolar coupling measurements

were carried out using liquid-crystalline Pf1 filamentous phages (Profos AG), as an alignment medium at a concentration of 7 mg/ml (residual quadrupolar coupling of ^2H_2O splitting = ca. 17.5 Hz), as described by Yang *et al.*[157]. To obtain an adequate amount of residual alignment, typically a ^2H_2O splitting of 10-20 Hz, is required.

The experimental residual dipolar couplings (RDCs) were correlated against the amino acids N93-K275 of the crystal structure of the *E.coli* maltose transporter (PDB code 2R6G)[45], employing the molecular alignment prediction program PALES[112]. Dipolar couplings were fitted using the best-fit module of the program. Three sets of dipolar couplings ($N_iH_i^N$, $C\alpha_iC'_i$, $H\alpha_iC\alpha_i$) were correlated to the amino acids (N93-K275), (N93-G260), (N93-S113, P209-T259) and (G117-T207) of MalF of the X-ray structure with data collected for the unbound state of MalF-P2 and one set of dipolar couplings ($N_iH_i^N$) were correlated to the amino acids (N93-K275), (N93-G260), (N93-S113, P209-T259) and (G117-T207) of MalF of the X-ray structure with data collected for the bound state of MalF-P2.
All experiments were zerofilled to 2048 points and 1024 points in the coupling resolved dimension in the 2D and 3D experiments respectively. Otherwise experiments were performed and analyzed as previous stated for standard 2D and 3D experiments. Dipolar couplings were extracted with CCPN[101] and further analyzed with Excel and Origin before analysis with Pales[112].

4.8.2 Computational methods

Methods used for the computer analysis and the MD simulations are described in the following two sections.

4.8.2.1 Molecular dynamics simulation of the MalF-P2 MalE complex

For every simulation ensemble, the MalE structure of the $MalFGK_2$-E crystal structure[45] and the structure of the P2-loop of MalF (residues F91-260) from the protein complex, a 50 ns molecular dynamic simulation in the GROMACS with the GROMOS force field (ffG53a6)[118] was performed. The proteins were placed into a triclinic box, which is 39% smaller than a rectangular system. The simulations were done under physiological conditions. For this purpose the proteins and the complex structure were calculated in a 0.1 M NaCl (e.g. ~15600 Spc H_2O molecules, 45 Na^+ and 33 Cl^- Ions, respectively) solution at 310 K and periodic boundary conditions for the box. Short range nonbonded interactions were calculated with the Cut-off Lennard-Jones potential up to a distance of 1.2 nm between the interacting atoms. For long range electrostatic interactions the particle mesh Ewald (PME) option was used with a grid spacing of 0.12 nm. The bond length of a heavy atom to an H-atom in protein or water was constrained using a LINCS and SETTLE algorithm, respectively. Every simulation ensemble was energy minimised in a two-step minimisation strategy. In a first step steepest descent and in a second a conjugate gradient minimising routine was used.
After equilibration over a period of 500 ps using a positional restraint of 1000 $kJmol^{-1}nm^{-2}$ on the backbone of the two proteins molecule, a free simulation was performed

for a period of 50 ns for the MalE-MalF-P2 complex. The temperature and pressure (1 bar) were controlled using the Berendsen coupling with a relaxation time of 0.1 ps for temperature and 1ps for pressure. During the course of the simulation, the actual frame was stored every 5 ps. Visualisation of trajectories and arrangement of the figures were realised using VMD[119].

4.8.2.2 Crystal structure analysis

Structure interpretations of the crystal structure of the maltose transporter[45] were done with the CCP4i. 1.3.20 (Computational Project, Number 4. 1994)[109] and PyMol[55]. Secondary structure elements of MalF-P2 in the X-ray structure were extracted using the dss algorithm in PyMol. Atom-atom distance search, between MalE and MalF-P2 in the crystal structure of the maltose transporter, were searched within the range 0.0 – 5.0 Å. All distances originate from heavy atoms N, C and O.

4.8.3 Isothermal Titration Calorimetry (ITC)

High-sensitivity microcalorimetry[158] was performed on a VP-ITC (MicroCal, Northampton, USA). Experiments were performed at 5 and 25 ^{0}C with MalE protein solutions in the calorimeter cell and MalF-P2 solutions in the injection syringe and in the reverse manner. The time spacings between the injections were chosen long enough to allow for complete re-equilibration (600 s). Typically, 20-40 injections were performed. Baseline subtraction and peak integration were accomplished using Origin 5.0 as described by the manufacturer (MicroCal Software, Northampton, USA). Curve fitting was performed with a non-linear fit based on a one-site binding model. The first injection was always excluded from evaluation because it usually suffers from sample losses due to mounting of the syringe and equilibration preceding the actual titration.

To prepare protein samples for ITC titration experiments, 2 ml of 1 mM solutions of MalE and MalF-P2 were dialyzed (Spectral/Por, Spectrum Laboratories Inc.) individually in 20 mM phosphate buffer (100 mM NaCl, pH 7.4) with and without maltose, concentrations ranging between 0.5-100 mM, over night at 7 ^{0}C with light stirring.

Blank experiments of MalF-P2 and MalE injected into pure buffer were performed, to measure the heat of dilution of the different protein solutions. Determination of binding of maltose to MalE was performed with MalE in the calorimetry cell (40 μM) and maltose in the injection syringe (500 μM). For MalF-P2 titrated into MalE, MalE (30-40 μM) and MalF-P2 (300-400 mM) were filtrated, degassed and inserted into the calorimeter cell and injection syringe, respectively. For the reverse performed titration MalF-P2 (50 mM) and MalE (600-900 mM) were filtrated, degassed and inserted into the calorimeter cell and injection syringe, respectively.

5. Figures and Tables

Figures

Tables

6. References

1. Jimonet, P. & Jager, R. Strategies for designing GPCR-focused libraries and screening sets. Current Opinion in Drug Discovery & Development 7, 325-333 (2004).
2. Kim, H., Melen, K., Osterberg, M. & von Heijne, G. A global topology map of the Saccharomyces cerevisiae membrane proteome. Proceedings of the National Academy of Sciences of the United States of America 103, 11142-11147 (2006).
3. Schnur, D. M., Hermsmeier, M. A. & Tebben, A. J. Are target-family-privileged substructures truly privileged? Journal of Medicinal Chemistry 49, 2000-2009 (2006).
4. Goffeau, A. et al. Life with 6000 genes. Science 274, 546-& (1996).
5. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Science 7, 1029-1038 (1998).
6. Hopkins, A. L. & Groom, C. R. The druggable genome. Nature Reviews Drug Discovery 1, 727-730 (2002).
7. Landry, Y. & Gies, J. P. Drugs and their molecular targets: an updated overview. Fundamental & Clinical Pharmacology 22, 1-18 (2008).
8. Myers, J. K., Beihoffer, L. A. & Sanders, C. R. Phenotology of disease linked proteins. Human Mutation 25, 90-97 (2005).
9. Sanders, C. R. & Myers, J. K. Disease-related misassembly of membrane proteins. Annual Review of Biophysics and Biomolecular Structure 33, 25-51 (2004).
10. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: Orientations of proteins in membranes database. Bioinformatics 22, 623-625 (2006).
11. Tusnady, G. E., Dosztanyi, Z. & Simon, I. Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics 20, 2964-2972 (2004).
12. Warschawski, D. The Warschawski lab. http://www.drorlist.com/nmr/MPNMR.html.
13. White, S.
14. Baldus, M. Magnetic resonance in the solid state: applications to protein folding, amyloid fibrils and membrane proteins. Eur Biophys J 36 Suppl 1, S37-48 (2007).
15. Hong, M. Structure, topology, and dynamics of membrane peptides and proteins from solid-state NMR Spectroscopy. Journal of Physical Chemistry B 111, 10340-10351 (2007).
16. Korukottu, J. et al. High-resolution 3D structure determination of kaliotoxin by solid-state NMR spectroscopy. PLoS One 3, e2359 (2008).
17. Li, Y., Berthold, D. A., Gennis, R. B. & Rienstra, C. M. Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy. Protein Science 17, 199-204 (2008).

18. Opella, S. J. & Marassi, F. M. Structure determination of membrane proteins by NMR spectroscopy. Chemical Reviews 104, 3587-3606 (2004).
19. Sanders, C. R. & Oxenoid, K. Customizing model membranes and samples for NMR spectroscopic studies of complex membrane proteins. Biochimica Et Biophysica Acta-Biomembranes 1508, 129-145 (2000).
20. Sanders, C. R. & Sonnichsen, F. Solution NMR of membrane proteins: practice and challenges. Magnetic Resonance in Chemistry 44, S24-S40 (2006).
21. Liang, B. Y. & Tamm, L. K. Structure of outer membrane protein G by solution NMR spectroscopy. Proceedings of the National Academy of Sciences of the United States of America 104, 16140-16145 (2007).
22. Jones, P. M. & George, A. M. The ABC transporter structure and mechanism: perspectives on recent research. Cellular and Molecular Life Sciences 61, 682-699 (2004).
23. Ponte-Sucre, A. The ABC Transporters in Microorganisms. (Caister Academic Press, 2009).
24. Chakraburtty, K. Translational regulation by ABC systems. Research in Microbiology 152, 391-399 (2001).
25. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiology and Molecular Biology Reviews 72, 317-364 (2008).
26. Goffeau, A., de Hertogh, B., Baret, P.V. ABC Transporters. In: Encyclopedia of Biological Chemistry. (Academic Press Inc, 2004).
27. Jones, P. M. & George, A. M. Multidrug resistance in parasites: ABC transporters, P-glycoproteins and molecular modelling. Int J Parasitol 35, 555-66 (2005).
28. Neijsel, O. M., Teixeira de Mattos, M.J., Tempest, D.W. Escherichia coli and Salmonell: Cellular and Molecular Biology (ed. Neidhardt, F. C.) (ASM, 1996).
29. Stouthamer, A. H. The search for correlation between theoretical and experimental growth yields. In: A theoretical evaluation of growth yields of yeasts. (Springer Netherlands, 1979).
30. Skou, J. C. The identification of the sodium-potassium pump (Nobel lecture). Angew. Chem. Int. Edn. Eng. 37, 2320-2328 (1998).
31. Dean, M., Rzhetsky, A. & Allikmets, R. The human ATP-binding cassette (ABC) transporter superfamily. Genome Research 11, 1156-1166 (2001).
32. Lage, L. ABC Transporters as target for RNA interference mediated ABC transporters in microorganisms. (Caister Academic Press, 2009).
33. Bennet, P. M. Plasmid encoded antibiotic resistance: acqusition and transfer of antiobiotic resistance genes in bacteria. (British Pharmacological Society, 2009).
34. Li, X. Z. & Nikaido, H. Efflux-mediated drug resistance in bacteria. Drugs 64, 159-204 (2004).
35. Ames, G. F. L. Bacterial Periplasmic Transport-Systems - Structure, Mechanism, and Evolution. Annual Review of Biochemistry 55, 397-425 (1986).

36. Doeven, M. K., Abele, R., Tampe, R. & Poolman, B. The binding specificity of OppA determines the selectivity of the oligopeptide ATP-binding cassette transporter. Journal of Biological Chemistry 279, 32301-32307 (2004).
37. Boos, W. & Shuman, H. Maltose/maltodextrin system of Escherichia coli: Transport, metabolism, and regulation. Microbiology and Molecular Biology Reviews 62, 204-+ (1998).
38. Zhang, Y. H. et al. Model of maltose-binding protein/chemoreceptor complex supports intrasubunit signaling mechanism. Proceedings of the National Academy of Sciences of the United States of America 96, 939-944 (1999).
39. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. Nature Reviews Molecular Cell Biology 10, 218-227 (2009).
40. Ehrmann, M., Ehrle, R., Hofmann, E., Boos, W. & Schlosser, A. The ABC maltose transporter. Molecular Microbiology 29, 685-694 (1998).
41. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly Related Sequences in the Alpha-Subunits and Beta-Subunits of Atp Synthase, Myosin, Kinases and Other Atp-Requiring Enzymes and a Common Nucleotide Binding Fold. Embo Journal 1, 945-951 (1982).
42. Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quiocho, F. A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. Molecular Cell 12, 651-661 (2003).
43. Dawson, R. J. P., Hollenstein, K. & Locher, K. P. Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. Molecular Microbiology 65, 250-257 (2007).
44. Hollenstein, K., Dawson, R. J. P. & Locher, K. P. Structure and mechanism of ABC transporter proteins. Current Opinion in Structural Biology 17, 412-418 (2007).
45. Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L. & Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. Nature 450, 515-U7 (2007).
46. Hvorup, R. N. et al. Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. Science 317, 1387-1390 (2007).
47. Daus, M. L., Berendt, S., Wuttge, S. & Schneider, E. Maltose binding protein (MalE) interacts with periplasmic loops P2 and P1 respectively of the MalFG subunits of the maltose ATP binding cassette transporter (MalFGK(2)) from Escherichia coli Salmonella during the transport cycle. Molecular Microbiology 66, 1107-1122 (2007).
48. Daus, M. L. et al. ATP induces conformational changes of periplasmic loop regions of the maltose ATP-binding cassette transporter. Journal of Biological Chemistry 281, 3856-3865 (2006).
49. Grote, M. et al. Transmembrane Signaling in the Maltose ABC Transporter MalFGK(2)-E PERIPLASMIC MalF-P2 LOOP COMMUNICATES SUBSTRATE AVAILABILITY TO THE ATP-BOUND MalK DIMER. Journal of Biological Chemistry 284, 17521-17526 (2009).
50. Kadaba, N. S., Kaiser, J. T., Johnson, E., Lee, A. & Rees, D. C. The high-affinity E-coli methionine ABC transporter: Structure and allosteric regulation. Science 321, 250-253 (2008).

51. Locher, K. P., Lee, A. T. & Rees, D. C. The E-coli BtuCD structure: A framework for ABC transporter architecture and mechanism. Science 296, 1091-1098 (2002).
52. Dawson, R. J. P. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. Nature 443, 180-185 (2006).
53. Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proceedings of the National Academy of Sciences of the United States of America 104, 19005-19010 (2007).
54. Hollenstein, K., Frei, D. C. & Locher, K. P. Structure of an ABC transporter in complex with its binding protein. Nature 446, 213-216 (2007).
55. Delano, W. L. ((San Carlos, CA, USA, 2002)).
56. Higgins, C. F. & Linton, K. J. The ATP switch model for ABC transporters. Nature Structural & Molecular Biology 11, 918-926 (2004).
57. Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P. & Rees, D. C. An inward-facing conformation of a putative metal-chelate-type ABC transporter. Science 315, 373-377 (2007).
58. Reyes, C. L., Ward, A., Yu, J. & Chang, G. The structures of MsbA: Insight into ABC transporter-mediated multidrug efflux. Febs Letters 580, 1042-1048 (2006).
59. Davidson, A. L. & Chen, J. ATP-binding cassette transporters in bacteria. Annual Review of Biochemistry 73, 241-268 (2004).
60. Locher, K. P. Structure and mechanism of ABC transporters. Current Opinion in Structural Biology 14, 426-431 (2004).
61. Oldham, M. L., Davidson, A. L. & Chen, J. Structural insights into ABC transporter mechanism. Current Opinion in Structural Biology 18, 726-733 (2008).
62. Pohl, A., Devaux, P. F. & Herrmann, A. Function of prokaryotic and eukaryotic ABC proteins in lipid transport. Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids 1733, 29-52 (2005).
63. Higgins, C. F. Abc Transporters - from Microorganisms to Man. Annual Review of Cell Biology 8, 67-113 (1992).
64. Senior, A. E., AlShawi, M. K. & Urbatsch, I. L. The catalytic cycle of P-glycoprotein. Febs Letters 377, 285-289 (1995).
65. Du, Z. M. et al. H-1, C-13, and N-15 NMR assignments of an engineered intein based on Mycobacterium tuberculosis RecA. Biomolecular Nmr Assignments 2, 111-113 (2008).
66. Mathys, S. et al. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene 231, 1-13 (1999).
67. Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. & Kyogoku, Y. Improved segmental isotope labeling of proteins and application to a larger protein. Journal of Biomolecular Nmr 14, 105-114 (1999).
68. Aller, S. G. et al. Structure of P-Glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. Science 323, 1718-1722 (2009).

69. Newstead, S. et al. Insights into How Nucleotide-Binding Domains Power ABC Transport. Structure 17, 1213-1222 (2009).
70. RCSB. The Resource to Protein Data Bank. http://www.rcsb.org.
71. Schwartz, M. Phenotypic expression and genetic localization of mutations affecting metabolism in Escherichia coli K 12. Ann Inst Pasteur 112, 673-698 (1967).
72. Salton M. J. R., K. K. S. Barons medical microbiology (ed. S., B.) (Univ of Texas Medical branch).
73. Ryan K. J., R. C. G. Medicl microbiology: an introduction to infectious disease. (McGraw-Hill Medical Publishing Division, New York, 2004).
74. Schirmer, T. & Cowan, S. W. Prediction of Membrane-Spanning Beta-Strands and Its Application to Maltoporin. Protein Science 2, 1361-1363 (1993).
75. Death, A. & Ferenci, T. Between Feast and Famine - Endogenous Inducer Synthesis in the Adaptation of Escherichia-Coli to Growth with Limiting Carbohydrates. Journal of Bacteriology 176, 5101-5107 (1994).
76. Notley, L. & Ferenci, T. Differential Expression of Mal Genes under Camp and Endogenous Inducer Control in Nutrient-Stressed Escherichia-Coli. Molecular Microbiology 16, 121-129 (1995).
77. Dutzler, R., Wang, Y. F., Rizkallah, P. J., Rosenbusch, J. P. & Schirmer, T. Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. Structure 4, 127-134 (1996).
78. Wang, Y. F., Dutzler, R., Rizkallah, P. J., Rosenbusch, J. P. & Schirmer, T. Channel specificity: Structural basis for sugar discrimination and differential flux rates in maltoporin. Journal of Molecular Biology 272, 56-63 (1997).
79. Boos, W. Periplasmic binding protein dependent ABC transporters. (American Society for Microbiology, Washington DC, 1996).
80. Davidson, A. L., Shuman, H. A. & Nikaido, H. Mechanism of Maltose Transport in Escherichia-Coli - Transmembrane Signaling by Periplasmic Binding-Proteins. Proceedings of the National Academy of Sciences of the United States of America 89, 2360-2364 (1992).
81. Dietzel, I., Kolb, V. & Boos, W. Pole Cap Formation in Escherichia-Coli Following Induction of Maltose-Binding Protein. Archives of Microbiology 118, 207-218 (1978).
82. Ferenci, T., Muir, M., Lee, K. S. & Maris, D. Substrate-Specificity of the Escherichia-Coli Maltodextrin Transport-System and Its Component Proteins. Biochimica Et Biophysica Acta 860, 44-50 (1986).
83. Lehmann, J., Schiltz, E. & Steck, J. Studies of the Interaction of the Maltose-Binding Protein of Escherichia-Coli, a Closed-Groove Binder, with 4,6-O-Ethylidenemalto-Oligosaccharides (Dp 2-5) and Its Regioselective Labeling with 3-Azibutyl 1-Thio-Alpha-(6-H-3) Maltoside. Carbohydrate Research 232, 77-87 (1992).
84. Martin, J. L., Waksman, G., Bardwell, J. C. A., Beckwith, J. & Kuriyan, J. Crystallization of Dsba, an Escherichia-Coli Protein Required for Disulfide Bond Formation Invivo. Journal of Molecular Biology 230, 1097-1100 (1993).
85. Freundlieb, S., Ehmann, U. & Boos, W. Facilitated Diffusion of Para-Nitrophenyl-Alpha-D-Maltohexaoside through the Outer-Membrane of

Escherichia-Coli - Characterization of Lamb as a Specific and Saturable Channel for Maltooligosaccharides. Journal of Biological Chemistry 263, 314-320 (1988).

86. Danot, O., VidalIngigliardi, D. & Raibaud, O. Two amino acid residues from the DNA-binding domain of MalT play a crucial role in transcriptional activation. Journal of Molecular Biology 262, 1-11 (1996).
87. Khare, D., Oldham, M. L., Orelle, C., Davidson, A. L. & Chen, J. Alternating Access in Maltose Transporter Mediated by Rigid-Body Rotations. Molecular Cell 33, 528-536 (2009).
88. Gould, A. D., Telmer, P. G. & Shilton, B. H. Stimulation of the Maltose Transporter ATPase by Unliganded Maltose Binding Protein. Biochemistry 48, 8051-8061 (2009).
89. MuroPastor, A. M., Ostrovsky, P. & Maloy, S. Regulation of gene expression by repressor localization: Biochemical evidence that membrane and DNA binding by the PutA protein are mutually exclusive. Journal of Bacteriology 179, 2788-2791 (1997).
90. Neuhard, J. Escherichia coli and Salmonella Typhimurioum: Cellular and Molecular Biology (ed. Ingrham J. L., N. F. C.) (ACM, Washington DC, 1987).
91. Saurin, W., Koster, W. & Dassa, E. Bacterial Binding Protein-Dependent Permeases - Characterization of Distinctive Signatures for Functionally Related Integral Cytoplasmic Membrane-Proteins. Molecular Microbiology 12, 993-1004 (1994).
92. Evenas, J. et al. Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy. Journal of Molecular Biology 309, 961-974 (2001).
93. Froshauer, S., Green, G. N., Boyd, D., McGovern, K. & Beckwith, J. Genetic analysis of the membrane insertion and topology of MalF, a cytoplasmic membrane protein of Escherichia coli. J Mol Biol 200, 501-11 (1988).
94. Daus, M. L. et al. ATP-driven MalK dimer closure and reopening and conformational changes of the "EAA" motifs are crucial for function of the maltose ATP-binding cassette transporter (MalFGK(2)) (vol 282, pg 22387, 2007). Journal of Biological Chemistry 282, 28296-28296 (2007).
95. Cabrita, L. D., Dai, W. W. & Bottomley, S. P. A family of E-coli expression vectors for laboratory scale and high throughput soluble protein production. Bmc Biotechnology 6, - (2006).
96. Czupalla, C. et al. Proteomic analysis of lysosomal acid hydrolases secreted by osteoclasts - Implications for lytic enzyme transport and bone metabolism. Molecular & Cellular Proteomics 5, 134-143 (2006).
97. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551-3567 (1999).
98. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological

macromolecules in solution. Proceedings of the National Academy of Sciences of the United States of America 94, 12366-12371 (1997).
99. Venters, R. A., Farmer, B. T., Fierke, C. A. & Spicer, L. D. Characterizing the use of perdeuteration in NMR studies of large proteins C-13, N-15 and H-1 assignments of human carbonic anhydrase II. Journal of Molecular Biology 264, 1101-1116 (1996).
100. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progress in Nuclear Magnetic Resonance Spectroscopy 34, 93-158 (1999).
101. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins-Structure Function and Bioinformatics 59, 687-696 (2005).
102. Wishart, D. S. & Sykes, B. D. Chemical-Shifts as a Tool for Structure Determination. Nuclear Magnetic Resonance, Pt C 239, 363-392 (1994).
103. Wishart, D. S., Sykes, B. D. & Richards, F. M. Relationship between Nuclear-Magnetic-Resonance Chemical-Shift and Protein Secondary Structure. Journal of Molecular Biology 222, 311-333 (1991).
104. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of Biomolecular Nmr 13, 289-302 (1999).
105. Daus, M. L. in Biologie Fakultet (Humboldt-Universität zu Berlin, Berlin, 2007).
106. Grote, M. in Biophysics (Humboldt-Universitat, Berlin, Berlin, 2008).
107. Jacso, T. et al. Periplasmic Loop P2 of the MalF Subunit of the Maltose ATP Binding Cassette Transporter Is Sufficient To Bind the Maltose Binding Protein MalE. Biochemistry 48, 2216-2225 (2009).
108. Daus, M. L., Grote, M. & Schneider, E. The MalF P2 Loop of the ATP-Binding Cassette Transporter MalFGK(2) from Escherichia coli and Salmonella enterica Serovar Typhimurium Interacts with Maltose Binding Protein (MalE) throughout the Catalytic Cycle. Journal of Bacteriology 191, 754-761 (2009).
109. Navaza, J. The CCP4 Suite: Programs for Protein Crystallography. Actra Cryst A50, 157-163 (1994).
110. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium (vol 278, pg 1111, 1997). Science 278, 1697-1697 (1997).
111. Prestegard, J. H., Al-Hashimi, H. M. & Tolman, J. R. NMR structures of biomolecules using field oriented media and residual dipolar couplings. Quarterly Reviews of Biophysics 33, 371-424 (2000).
112. Zweckstetter, M. & Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. Journal of the American Chemical Society 122, 3791-3792 (2000).
113. Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nature Structural Biology 5, 1065-1074 (1998).

114. Hansen, M. R., Hanson, P. & Pardi, A. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. Rna-Ligand Interactions Pt A 317, 220-240 (2000).
115. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. Journal of Magnetic Resonance 160, 65-73 (2003).
116. Hus, J. C. et al. 16-Fold Degeneracy of Peptide Plane Orientations from Residual Dipolar Couplings: Analytical Treatment and Implications for Protein Structure Determination. Journal of the American Chemical Society 130, 15927-15937 (2008).
117. Bax, A. & Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. Current Opinion in Structural Biology 15, 563-570 (2005).
118. Van der Spoel, D. et al. Gromacs: Fast, Flexible, and Free. Journal of Computational Chemistry 26, 1701-1718 (2005).
119. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. Journal of Molecular Graphics 14, 33-& (1996).
120. Gryk, M. R., Abseher, R., Simon, B., Nilges, M. & Oschkinat, H. Heteronuclear relaxation study of the PH domain of beta-spectrin: Restriction of loop motions upon binding inositol trisphosphate. Journal of Molecular Biology 280, 879-896 (1998).
121. Palmer, A. G., Williams, J. & McDermott, A. Nuclear magnetic resonance studies of biopolymer dynamics. Journal of Physical Chemistry 100, 13293-13310 (1996).
122. Kay, L. E., Muhandiram, D. R., Wolf, G., Shoelson, S. E. & Forman-Kay, J. D. Correlation between binding and dynamics at SH2 domain interfaces. Nature Structural Biology 5, 156-163 (1998).
123. Bracken, C., Carr, P. A., Cavanagh, J. & Palmer, A. G. Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: Implications for the entropy of association with DNA. Journal of Molecular Biology 285, 2133-2146 (1999).
124. Kovrigin, E. L., Kempf, J. G., Grey, M. J. & Loria, J. P. Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. Journal of Magnetic Resonance 180, 93-104 (2006).
125. Loria, J. P., Rance, M. & Palmer, A. G. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. Journal of the American Chemical Society 121, 2331-2332 (1999).
126. Martin, M. L., Martin, G. J., Delpeuch, J. J. Practical NMR Spectroscopy (AAA, 1980).
127. Davis, D. G., Perlman, M. E. & London, R. E. Direct Measurements of the Dissociation-Rate Constant for Inhibitor-Enzyme Complexes Via the T-1-Rho and T-2 (Cpmg) Methods. Journal of Magnetic Resonance Series B 104, 266-275 (1994).

128. Phan, I. Q. H., Boyd, J. & Campbell, I. D. Dynamic studies of a fibronectin type I module pair at three frequencies: Anisotropic modelling and direct determination of conformational exchange. Journal of Biomolecular Nmr 8, 369-378 (1996).
129. Peng, J. W. & Wagner, G. Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. Biochemistry 34, 16733-16752 (1995).
130. Ishima, R. & Torchia, D. A. Estimating the time scale of chemical exchange of proteins from measurements of transverse relaxation rates in solution. Journal of Biomolecular Nmr 14, 369-372 (1999).
131. McNaught, A. D. IUPAC: Compendium of Chemical Terminology (ed. A., J.) (Blackwell Scientific Publications, Oxford, 2006).
132. Millet, O., Loria, J. P., Kroenke, C. D., Pons, M. & Palmer, A. G. The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. Journal of the American Chemical Society 122, 2867-2877 (2000).
133. KAY, L. E., Nicholson, L. K., Delaglio, F., Bax, A. & Torchia, D. A. Pulse Sequences for Removal of the Effects of Cross-Correlation between Dipolar and Chemical-Shift Anisotropy Relaxation Mechanism on the Measurement of Heteronuclear T1 and T2 Values in Proteins. Journal of Magnetic Resonance 97, 359-375 (1992).
134. Tapia, M. I., Mourez, M., Hofnung, M. & Dassa, E. Structure-function study of MalF protein by random mutagenesis. J Bacteriol 181, 2267-72 (1999).
135. Merino, G., Boos, W., Shuman, H. A. & Bohl, E. The inhibition of maltose transport by the unliganded form of the maltose-binding protein of Escherichia coli: experimental findings and mathematical treatment. J Theor Biol 177, 171-9 (1995).
136. Austermuhle, M. I., Hall, J. A., Klug, C. S. & Davidson, A. L. Maltose-binding protein is open in the catalytic transition state for ATP hydrolysis during maltose transport. J Biol Chem 279, 28243-50 (2004).
137. Levitt, M. H. Spin Dynamics: Basics of Nuclear Magnetic Resonance (Wiley, 2008).
138. Matsuo, H. et al. Identification by NMR spectroscopy of residues at contact surfaces in large, slowly exchanging macromolecular complexes. Journal of the American Chemical Society 121, 9903-9904 (1999).
139. Korzhnev, D. M. et al. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. Nature 430, 586-590 (2004).
140. Shilton, B. H. The dynamics of the MBP-MalFGK(2) interaction: A prototype for binding protein dependent ABC-transporter systems. Biochimica Et Biophysica Acta-Biomembranes 1778, 1772-1780 (2008).
141. Ehrmann, M. & Beckwith, J. Proper Insertion of a Complex Membrane-Protein in the Absence of Its Amino-Terminal Export Signal. Journal of Biological Chemistry 266, 16530-16533 (1991).
142. Holland, B., Cole, S. P., Kuchler, K. & Higgins, C. F. ABC Proteins: From Bacteria to Man (Academic Press, new York, 2002).

143. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 449, 1078-U12 (2007).
144. HOR, L. I. & Shuman, H. A. Genetic-Analysis of Periplasmic Binding-Protein Dependent Transport in Escherichia-Coli - Each Lobe of Maltose-Binding Protein Interacts with a Different Subunit of the Malfgk(2) Membrane-Transport Complex. Journal of Molecular Biology 233, 659-670 (1993).
145. Millet, O., Hudson, R. P. & Kay, L. E. The energetic cost of domain reorientation in maltose-binding protein as studied by NMR and fluorescence spectroscopy. Proceedings of the National Academy of Sciences of the United States of America 100, 12700-12705 (2003).
146. Telmer, P. G. & Shilton, B. H. Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. Journal of Biological Chemistry 278, 34555-34567 (2003).155. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. Journal of Magnetic Resonance 131, 373-378 (1998).
147. Sambrook, J. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 2001).
148. Gasteiger, E. et al. The Proteomics Protocols Handbook (Humana Press, 2005).
149. Schuck, P. & Rossmanith, P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. Biopolymers 54, 328-41 (2000).
150. Yoshimura, Y. et al. Identification of protein substrates of Ca^{2+}/calmodulin-dependent protein kinase II in the postsynaptic density by protein sequencing and mass spectrometry. Biochemical and Biophysical Research Communications 290, 948-954 (2002).156. Bax, A., Kontaxis, G. & Tjandra, N. Dipolar couplings in macromolecular structure determination. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B 339, 127-174 (2001).
151. Loo, T. W. & Clarke, D. M. Determining the dimensions of the drug-binding domain of human P-glycoprotein using thiol cross-linking compounds as molecular rulers. Journal of Biological Chemistry 276, 36877-36880 (2001).
152. Farrow, N. A. et al. Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by N-15 Nmr Relaxation. Biochemistry 33, 5984-6003 (1994).
153. Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. Journal of the American Chemical Society 124, 12352-12360 (2002).
154. Tollinger, M., Skrynnikov, N. R., Mulder, F. A. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. Journal of the American Chemical Society 123, 11341-11352 (2001).

155. Bax, A., Kontaxis, G. & Tjandra, N. Dipolar couplings in macromolecular structure determination. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B 339, 127-174 (2001).
156. Koenig, B. W. et al. Measurement of dipolar couplings in a transducin peptide fragment weakly bound to oriented photo-activated rhodopsin. Journal of Biomolecular Nmr 16, 121-125 (2000).
157. Yang, D. W., Venters, R. A., Mueller, G. A., Choy, W. Y. & Kay, L. E. TROSY-based HNCO pulse sequences for the measurement of (HN)-H-1-N-15, N-15-(CO)-C-13, (HN)-H-1-(CO)-C-13, (CO)-C-13-C-13(alpha) and (HN)-H-1-C-13(alpha) dipolar couplings in N-15, C-13, H-2-labeled proteins. Journal of Biomolecular Nmr 14, 333-343 (1999).
158. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. Analytical Biochemistry 179, 131-137 (1989).

Appendix

A.1 NMR

The explosive growth in the field of nuclear magnetic resonance (NMR) spectroscopy that continues today originated with the development of pulsed Fourier transform NMR spectroscopy by Ernst and Anderson[1] and the conception of multidimensional NMR spectroscopy by Jeener[2]. Nuclear magnetisation resonances in bulk condensed phase were reported for the first time in 1946 by Bloch *et al*.[3] and by Purcell *et al*.[4]. Currently, NMR spectroscopy and X-ray crystallography are the only techniques capable of determining the three-dimensional structures of macromolecules at atomic resolution. In addition, NMR spectroscopy is a powerful technique for investigating time-dependent chemical phenomena, including reaction kinetics and intramolecular dynamics.
Whereas X-ray analysis is based on diffraction patterns of crystals yielding electron densities, NMR is characterized by dynamic properties, or rather the quantum mechanical properties of nuclear spin angular momentum. The origins of these physical properties have been discussed in the review articles of Sloan *et al*.[5] and Rith *et al*.[6]

Currently most of the recent innovations within NMR spectroscopy have been within the field of protein NMR, which has become a very important technique in structural biology. In contrast to X-ray crystallography, NMR is primarily limited to relatively small proteins, usually smaller than 35 kDa, though technical advances allow ever-larger structures to be solved. Also, NMR spectroscopy is often the only way to obtain high resolution information on partially or wholly intrinsically unstructured proteins, which don't crystallize[7-10]. This class of proteins have been shown to be extremely important for a range of actions within the cell[8].
Because of the high amount of atoms in macromolecules, basic 1D NMR techniques cannot be used due to spectral overlap making the analysis impossible. Because of this 2D, 3D and 4D (and even higher dimensions) type of experiments have been devised to resolve this problem. To be able to employ these experiments to their full content, biochemistry has come to aid, by the addition of being able to add more NMR active nuclei in different affordable fashions. Proteins can now be isotopically labelled with their NMR active nuclei ^{13}C and ^{15}N, uniformly or even residue specific.

Due to the complexity and vast amount of easily attainable articles regarding quantum mechanical phenomena in NMR and for bimolecular applications[11-14], only topics of interest to this work will be described. This chapter will instead emphasize more on the practical side of protein NMR analysis and cover some of the most useful structural biology aspects of NMR. These include chemical shift assignments, structural properties linked to chemical shift analysis, ligand-protein interactions (protein-protein interactions apply to the same theory), protein dynamics and angular restraints in the form of residual dipolar couplings (RDCs). My wish with this is to give a basic understanding on how to approach a structural and functional study of a macromolecule with protein NMR spectroscopy as the major tool.

A.1.1 NMR Theory

The origin of NMR spectroscopy is based on the spin angular momentum of the different chemical elements. The spin angular momentum is characterized by the nuclear spin quantum number, I. Although NMR takes the nuclear spin as a given quantity, certain systematic features can be noted: (i) nuclei with odd mass numbers have half-integral spin quantum numbers, (ii) nuclei with an even mass number and an even atomic number have spin quantum numbers equal to zero, and (iii) nuclei with an even mass number and an odd atomic number have integral spin quantum numbers. Because the NMR phenomenon relies on the nuclear spin, atoms of category (ii) are NMR inactive. Nuclei with spin numbers greater than ½ also possess electric quadrupole moments arising from non-spherical nuclear charge distributions. The lifetimes of the magnetic states for quadrupolar nuclei in solution normally are much shorter than are the lifetimes for nuclei with I = ½. NMR resonances for quadrupolar nuclei are correspondingly broad and can be more difficult to study. Relevant properties of nuclei commonly found in biomolecules are given in Table A.1.1.1. The most important nuclei for biomolecular NMR with I = ½ are ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P; the most important nuclei with I = 1 is ^{2}H. Most chemical elements have one or more stable isotopes with I = ½, though these might be of very low abundance and can be extremely costly to incorporate, i.e. ^{17}O has I = 5/2 but with a natural abundance of 0.037% and a cost of ~1250 $/g liquid.

Table A.1.1.1. Properties of selected nuclei. Shown are the chemical elements of particular use in biomolecular NMR, with their spin angular momentum quantum number, I, the magnetogyrotic ratio, γ, and their natural isotopic abundance.

Nucleus	*I*	$\gamma(\mathrm{Ts})^{-1}$	Natural Abundance (%)
^{1}H	1/2	2.6753×10^{8}	99.99
^{2}H	1	4.107×10^{7}	0.012
^{13}C	1/2	6.728×10^{7}	1.07
^{14}N	1	1.934×10^{7}	99.63
^{15}N	1/2	-2.713×10^{7}	0.37
^{17}O	5/2	-3.628×10^{7}	0.038
^{19}F	1/2	2.518×10^{8}	100
^{31}P	1/2	1.0839×10^{8}	100
^{113}Cd	1/2	-5.961×10^{7}	12.22

The nuclear spin angular momentum, **I**, is a vector quantity with magnitude given by

$$[\mathbf{I}] = [\mathbf{I} \bullet \mathbf{I}]^{1/2} = \hbar[I(I+1)]^{1/2} \qquad [A.1.1.1]$$

in which I is the nuclear spin angular momentum quantum number and $\hbar$ is Planck's constant divided by 2π. Due to the restrictions of quantum mechanics, only one the three Cartesian components of **I** can be specified simultaneously with $\mathbf{I}^2 \equiv \mathbf{I} \bullet \mathbf{I}$. By

convention, the value of the z-component of **I** is specified by the following equation:

$$I_z = \hbar m \qquad \text{[A.1.1.2]}$$

in which the magnetic quantum number m = (-I, -I+1, …, I-1, I). Thus, I_z has 2I+1 possible values. The orientation of the spin angular momentum vector in space is quantized, because the magnitude of the vector is constant and the z-component has a set of discrete possible values. In the absence of external fields, the quantum states corresponding to the 2I+1 values of m have the same energy, and the spin angular momentum vector does not have a preferred orientation.
Nuclei that have nonzero spin angular momentum also possess nuclear magnetic moments. As a consequence of the Wigner-Eckart theorem[15], the nuclear magnetic moment, **μ**, is collinear with the vector representing the nuclear spin angular momentum vector and is defined by

$$\boldsymbol{\mu} = \gamma \mathbf{I} \qquad \text{[A.1.1.3]}$$

$$\mu_z = \gamma I_z = \gamma \hbar m \qquad \text{[A.1.1.4]}$$

in which the magnetogyric ratio, γ, is a characteristic constant for a given nucleus (Table A.1.1.1). Because the angular momentum is a quantized property, the nuclear magnetic moment is quantized. This can graphically be presented as a gyroscope that has an angular momentum that is fixed in three-dimensional space (Figure A.1.1.1),

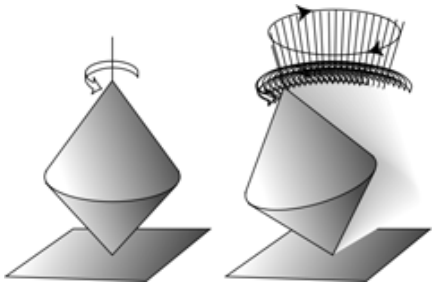

Figure A.1.1.1. The spin angular momentum of a nucleus represented as a gyroscope whose axis is fixed in three-dimensional space.

where the magnitude of the γ, in part, determines the receptivity of a nucleus in NMR spectroscopy. When the nucleus is affected by an external magnetic field it will start precess about it. The angular frequency at which this precession occurs is given by

$$\upsilon = \frac{\gamma H}{2\pi} \qquad \text{[A.1.1.5]}$$

where γ is the gyromagnetic ratio of the nucleus, H the strength of the applied external magnetic field, and ν is the angular frequency, called the Larmor frequency. For

example, a ^{1}H atom in presence of a magnetic field, H = 14.1T, has a Larmor frequency of 600MHz. The spin states energies of the nucleus are given by

$$E = -\boldsymbol{\mu} \bullet \mathbf{B} \qquad \text{[A.1.1.6]}$$

in which **B** is the magnetic field vector. The minimum energy is obtained when the projection of μ onto **B** is maximized. Because $|\mathbf{I}| > \mathbf{I}_z$, with energies proportional to their projection onto **B**. in an NMR spectrometer, the static external magnetic field is directed by convention along the z-axis of the laboratory coordinate system. For this geometry, Eq. [A.1.1.6] reduces to

$$E_m = -\gamma I_z B_0 = -m\hbar\gamma B_0 \qquad \text{[A.1.1.7]}$$

in which B_0 is the static magnetic field strength. In the presence of a static magnetic field, the projections of the angular momentum of the nuclei onto the z-axis of the laboratory frame results in 2I+1 equally spaced energy levels, which are known as Zeeman levels. The optimization of I_z is illustrated in Figure A.1.1.2.

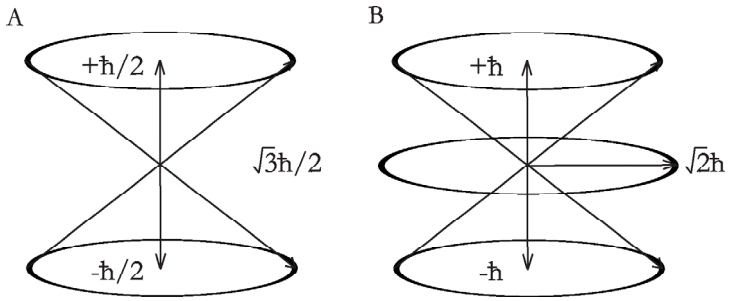

Figure A.1.1.2. Angular momentum vectors, **I**, and the allowed z-components, I_z, for panel A a spin ½ particle and panel B a spin 1 particle. The location of **I** on the surface of the cone cannot be specified because of quantum mechanical uncertainties in the I_x and I_y components.

At equilibrium the different energy states are unequally populated because lower energy orientations of the magnetic dipole are more probable. The relative population of a state is given by the Boltzmann distribution,

$$\frac{N_{upper}}{N_{lower}} = e^{-\gamma_N Hh/kT} \qquad \text{[A.1.1.8]}$$

where N_{upper} and N_{lower} are the numbers of nuclei in the respective spin states, γ is the gyromagnetic ratio, h is Planck's constant, H is the external magnetic field, k is the Boltzmann constant, and T is the temperature. For ^{1}H at room temperature subjected to a 600MHz external magnetic field, the ratio is 0.9999382. The populations of the states depend both on the nucleus type and on the applied magnetic field, as the external magnetic field strength increases the nuclear spin energy levels become larger and the population between the states increase. The sensitivity of NMR depends upon these population differences and as a result NMR is a very insensitivity

method in comparison to other spectroscopic techniques and has lead to a great drive to construct more powerful magnets. For an isolated spin, with I = ½, only two nuclear spin states exist and two energy levels separated by $\Delta E = \hbar\gamma B_0$ are obtained by application of an external magnetic field. A single Zeeman transition between the energy levels exists. The spin state with m = + ½ is referred to as the α state, and the state with m = - ½ is referred to as the β state. If γ is positive (negative), then the α state has lower (higher) energy compared to the β state. This is illustrated in Figure A.1.1.3.

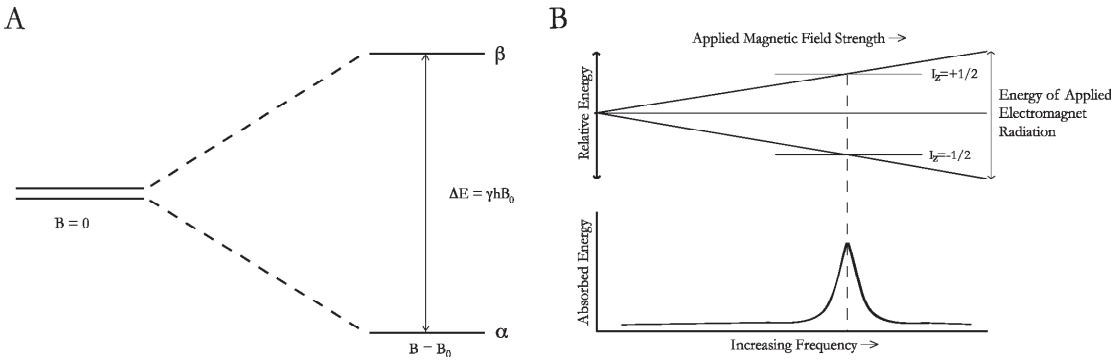

Figure A.1.1.3. The increased energy levels, when an external magnetic field is applied, are depicted in panels A and B. The energy level increase between α- and β-states in relation to increase in applied external magnet field and its corresponding frequency change are shown in b).

As a consequence of the angular momentum of the nuclei affected by the B_0 field, the nuclei will have a specific precession frequency, given by its specific Zeeman energy levels. In macromolecules the different nuclei will not only be affected by the B_0, but also from the local electro magnetic fields from other atoms. The perturbation of the B_0 field due to local magnetic fields changes the Zeeman levels slightly and hence the frequency of the nuclei. This is often a subtle change but is also responsible for giving the very detailed information given by NMR. This very important effect is called the *chemical shift* and will be discussed further in the next sections.

Up to now only a simple system of a homogenous sample of non-interacting nuclei has been considered. In macromolecular samples the amino acids themselves, and them being incorporated into long polypeptide chains, gives complex systems with different nuclei coupled to one another by chemical bonds. Nuclei interact with each other either by dipolar couplings or through scalar couplings often named J-couplings. High resolution NMR is able to reveal fine structure due to these interactions between the nuclei, also recall the *chemical shift* effect. The interaction between nuclei result in a splitting of the resonance signals into multiplets, these splittings are not caused by the direct dipolar interactions between magnetic dipole moments. The dipolar coupling is an anisotropic quantity that is averaged to zero to first order in isotropic solution (though this effect is extremely important in the anisotropic case in solid state NMR). Ramsey and Purcell suggested that the interaction is rather mediated by the electrons forming the chemical bonds between the nuclei[16]. This interaction is known as the spin-spin coupling or scalar coupling. The strength of the interaction is

measured by the scalar coupling constant, $^{n}J_{ab}$, in which n designates the number of covalent bonds separating the two nuclei, a and b. The magnitude of $^{n}J_{ab}$ is usually expressed in Hertz (Hz) and the most important scalar coupling interactions in proteins have n = 1 to 4. In the present text, n will be written explicitly only if the intended value of n is not clear from the context.
Scalar couplings modify the energy levels of the system, and the NMR spectrum is modified correspondingly. The prototypical example consists of two spin ½ nuclei, most often of $^{1}H - ^{1}H$ or $^{1}H - ^{13}C$ spins. The two spins are designated I and S. The resonance frequencies are ω_I and ω_S, respectively,

$$\omega_I = -\gamma B_0 (1 - \sigma_I) \qquad \text{[A.1.1.9]}$$

The magnetic quantum numbers are m_I and m_S; each spin has two stationary states that correspond to the magnetic quantum numbers ½ and –½. The complete two-spin system is described by four wave functions corresponding to all possible combinations of m_I and m_S,

$$\psi_1 = \psi\left(\frac{1}{2},\frac{1}{2}\right) \qquad \psi_2 = \psi\left(\frac{1}{2},-\frac{1}{2}\right)$$
$$\psi_3 = \psi\left(-\frac{1}{2},\frac{1}{2}\right) \qquad \psi_4 = \psi\left(-\frac{1}{2},-\frac{1}{2}\right) \qquad \text{[A.1.1.10]}$$

where the first quantum number describes the state of the I spin and the second describes the S spin. In the absence of scalar coupling between the spins, the energies of these four states are the sums of the energies for each spin. Remembering that the β state has higher (lower) energy compared to the α state for positive (negative) γ, the energies are found to be

$$E_1 = \frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S \qquad E_2 = \frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S$$
$$E_3 = -\frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S \qquad E_4 = -\frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S \qquad \text{[A.1.1.11]}$$

The total magnetic quantum number m for each energy level is the sum of the individual terms

$$\omega_{24} = \omega_I - \pi J_{IS} \qquad m_2 = \frac{1}{2} - \frac{1}{2} = 0$$
$$m_3 = -\frac{1}{2} + \frac{1}{2} = 0 \qquad m_4 = -\frac{1}{2} - \frac{1}{2} = 0 \qquad \text{[A.1.1.12]}$$

The energy level diagram for a two-spin system with $\gamma_I > \gamma_S > 0$ is shown in Figure A.1.1.4 a). The observable transitions obey the selection rule $\Delta m = \pm 1$. Therefore, the allowed transitions occur between states 1-2, 3-4, 1-3 and 2-4, Figure A.1.1.4 a) and b); transitions between 2-3 or 1-4 are forbidden. The first two transitions involve a change in the spin state of the S spin while the latter two involve a change in the spin state of the I spin. Consequently, the NMR spectrum shown in Figure A.1.1.4 c) consists of one resonance line at ω_I, due to transitions 1-3 and 2-4, and one resonance line at ω_S, due to transitions 1-2 and 3-4.

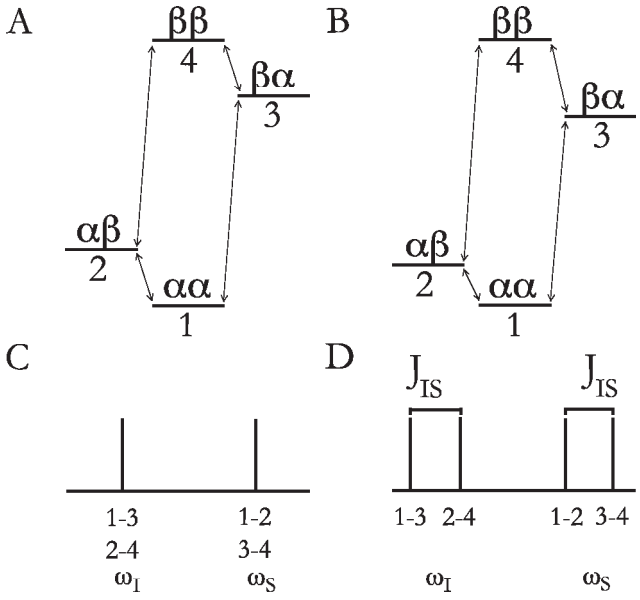

Figure A.1.1.4. Energy levels for an AX spin system. Shown are the energy levels for an AX spin system in the a) absence and b) presence of scalar coupling interactions between the spins, assuming $J_{IS} > 0$ and $\gamma_I > \gamma_S > 0$. The allowed transitions are indicated between arrows. The energies of the four spin states are defined by the equations; a) [A.1.1.11] and b) [A.1.1.13].

Introducing the scalar coupling between I and S, with a value of J_{IS}, modifies the energy levels to

$$E_1 = \frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S + \frac{1}{2}\pi\hbar J_{IS} \qquad E_2 = \frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S - \frac{1}{2}\pi\hbar J_{IS}$$

$$E_3 = -\frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S - \frac{1}{2}\pi\hbar J_{IS} \qquad E_4 = -\frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S + \frac{1}{2}\pi\hbar J_{IS} \qquad \text{[A.1.1.13]}$$

in which weak coupling has been assumed with $2\pi|J_{IS}| \ll |\omega_I - \omega_S|$. These expressions are derived from

$$E(m_I, m_S) = m_I\omega_I + m_S\omega_S + 2\pi m_I m_S J_{IS} \qquad \text{[A.1.1.14]}$$

The term in J_{IS} depends on the spin states of both nuclei but the terms in ω_I and ω_S depend on the spin state of a single nucleus. The energy level diagram for a scalar coupled twp-spin system is shown in Figure A.1.1.4 b), assuming that $J_{IS} > 0$. The resulting effect in the spectrum of the scalar coupled system is easily seen from the new values from the transition frequencies,

$$\omega_{12} = \omega_S + \pi J_{IS} \qquad \omega_{34} = \omega_S - \pi J_{IS}$$
$$\omega_{13} = \omega_I + \pi J_{IS} \qquad \omega_{24} = \omega_I - \pi J_{IS} \qquad \text{[A.1.1.15]}$$

Now the spectrum shown in Figure A.1.1.4 d) consists of four lines: two centred around the transition frequency, ω_S, of the S spin but separated by $2\pi J_{IS}$, and two centred around the transition frequency of the I spin, ω_I, but separated by $2\pi J_{IS}$. A weakly coupled two-spin system is referred to as an AX spin system and a strongly coupled two-spin system is referred to as an AB spin system, in which A and X or A and B represent the pair of scalar coupled spins.
To overcome spectral complexity, due to the peak splitting in the spectra, a given set of nuclei is irradiated with a strong rf energy at its characteristic Larmor frequency. This causes spin saturation and rapid interconversion of the spin states. Neighbouring nuclei with different Larmor frequencies are no longer influenced by specific long-lived spins, so spin-spin signal splitting of the neighbouring nuclei vanishes. The effect is shown for a 2D spectrum in Figure A.1.1.5. Decoupling can be applied either to the same type of nuclei, homonuclear decoupling, or of different nuclei, heteronuclear decoupling. Different methods of decoupling are well described by Shaka, A., J., in the Encyclopaedia of Nuclear Magnetic Resonance, chapter of decoupling methods[12]. Other methods of decoupling can also be found from the group of Glaser *et al.*[17-19].

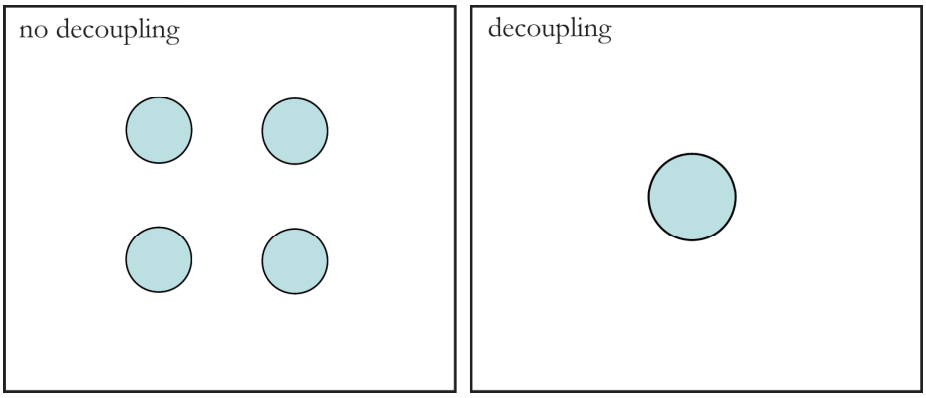

Figure A.1.1.5. Upon decoupling, a strong rf field at a specific nuclei's Larmor frequency, the peak splitting collapses into one peak, due to the spin saturation and rapid interconversion of the spin states. The effect of decoupling is here shown in a 2D spectrum, taking the J-coupling splitting effect away, greatly reducing the complexity.

In most cases it is beneficial to use decoupling when acquiring a spectrum, but there

are also cases when this do not apply. Even though decoupling is very useful the high *rf* field applied can cause severe heating to the sample, and this can be a problem for sensitive protein samples.
The next coming sections will further address some of the already mentioned effects and principles together with new ones.

A.1.2 Chemical shift assignments

The most time consuming step of the interpretation of NMR data is to assign resonances to specific atoms. NMR can readily give a vast amount of information of the studied system, but only if one is be able to identify it. Here, one of the earlier mentioned principles, the *chemical shift*, will be explained further.
The chemical shift describes the dependence of nuclear magnetic energy levels on the electric environment in a molecule. The net field of the nuclei will not only depend on the external magnetic field, B_0, but also by the local magnetic fields induced by currents of electrons in the molecular orbitals of the nuclei. The effect of these secondary fields is called nuclear shielding and can augment or diminish the effect of the main field. Fluctuations in the local magnetic field as the molecule rotate in solution also results in the chemical shift anisotropy, CSA These fluctuations cause the spin energy levels to change (and their resonance frequencies). The variations of nuclear magnetic resonance frequencies of the same kind of nuclei, is called the chemical shift. The size of the chemical shift is given with respect to a reference frequency or a reference sample, usually a molecule with barely distorted electron distribution; the most common molecule used for this is TMS, tetramethylsilane. The detected ^{1}H, ^{13}C chemical shifts of a molecule are referenced to the ^{1}H, ^{13}C chemical shifts of TMS on the ppm scale. Other molecules are used for setting the chemical shift of other nuclei.
The chemical shift δ is usually expressed in parts per million (ppm or δ) by frequency, given by

$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \times 10^6 \qquad \text{[A.1.2.1]}$$

in which Ω and Ω_{ref} are the offset frequencies of the signal of interest and the reference signal, respectively. ω_0 is the applied static magnetic field.
Some of the most important factors causing chemical shifts are; a nucleus in the vicinity of an electronegative atom, which reduces the electron density around it and deshields the nucleus. Anisotropic induced magnetic fields is a result of a locally induced magnetic field experienced by nucleus resulting from circulating electrons that can be either paramagnetic when it is parallel to the applied magnetic field or diamagnetic when it is opposed to it. It is observed in alkenes and alkynes. In alkenes the double bond is oriented perpendicular to the external field with π electrons likewise circulating at right angles. The induced magnetic field lines are parallel to the external field at the location of the alkene protons and deshield the ^{1}H nuclei. The

three-dimensional space where a nucleus experiences diamagnetic shift is called the shielding zone with a cone shape aligned to the external field. This is even more prominent in aromatic compounds where the aromatic ring creates a diamagnetic ring current, typically protons in aromatic compounds resonance at 7.5 ppm.
In alkynes the most effective orientation is the external field in parallel with the electrons circulation around the triple bond. As a result the acetylenic protons are located in the cone shaped shielding zone and resonance upfield, in contrast to the alkenes. Figure A.1.2.1 illustrates the alkene and alkyne anisotropic chemical shift effects.

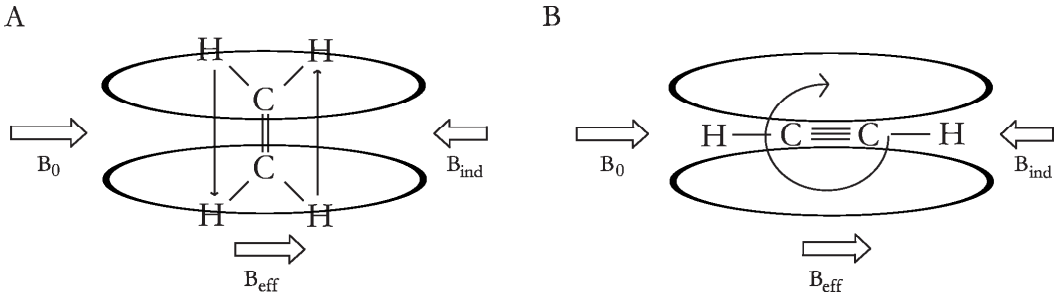

Figure A.1.2.1 Shown in ellipses are the three dimensional space where a nucleus experiences a diamagnetic shift for alkenes, a), and alkynes, b).

In protein NMR one can divide the full chemical shift analysis into two parts, backbone- and sidechain-assignments of all the individual amino acids present in the macromolecule.
But, before chemical shift analysis should be started one has to make sure that the protein in question is suitable for NMR. To be able to perform the necessary 3D experiments (1D and 2D experiments are in theory enough but with larger systems > 5 kDa the analysis will be enormously complex and not manageable) for chemical shift analysis the sample has to fulfil some requirements, (i) have a well purified sample with concentrations ranging between 0.7 – 2.0 mM of isotopically labelled protein, (ii) the protein should be within the NMR size limits 5 – 40 kDa (as of 2008 the biggest solved monomer protein structure solved by NMR is of the maltose binding protein (MBP) at 41.2 kDa, pdb code 2v93) to have better relaxation properties and also (iii) give a well resolved spectrum. There are ways to circumvent the spectral overlapping in larger systems, but they often require extended biochemical work, i.e. residue specific labelling can be used to reduce the complexity by only make a few residue types NMR active[24-26] and/or intein technology to only make specific smaller domains of a large protein NMR active[27, 28].

The first part to be considered is the backbone-assignments, which comprises all the atoms in the main chain of the amino acids, from which the various side chains project. These include H^N and N from the amino group, CO from the carboxyl group and Cα, Cβ (although Cβ is not actually present in the main chain).
A strategy for backbone assignment will be presented in A.1.2.1 Backbone

assignments. The second part considers the assignments of the side chains of the amino acids and one soon understands why 3D and 4D types of experiments were implemented to make the assignments more approachable. Side chain assignments are the most time consuming steps but are vital to high resolution structure determination. This will be presented in A.1.2.2 Side chain assignments.

Before one starts a NMR project one should carefully consider what the goal of it is, because many times only a main chain assignment is needed to fulfill the goal/s and much time can be saved leaving out the time consuming side chain assignments. Due to this consuming work of assignments some journals have specialised in only documenting this, journals as Journal of Biomolecular NMR Assignments, and also the BioMagResBank (BMRB) was established, which contain a large databank with deposited chemical shifts, relaxation data and structural data for a large range of macromolecules and peptides.

In protein NMR there are many ways that one can approach assignments of chemical shifts, due to different experiments recorded and even software available at hand, and I will only address the way that was used in this work.

A.1.2.1 Backbone assignments

The most common method of backbone resonance assignment involves the use of uniformly ^{15}N, ^{13}C labelled protein and the measurement of CBCANNH and CBCA(CO)NNH spectra.
For smaller proteins ranging between 5 – 20 kDa the so called CBCA pair experiments are sufficient to get the HN, N, Cα, Cβ assignments. Proteins of this size have a τ_c value typically less than 12 ns[29], and therefore have relaxation properties well suited for these experiments. For larger proteins that does not yield spectra of sufficient quality, complimentary experiments as HNCA, HN(CO)CA, HNCO and HN(CA)CO can greatly help. These experiments are more sensitive and also have a reduced amount of resonances. How these experiments are recorded is found in chapter 4.9.1.1.
For very large proteins, > 20 kDa, it may become necessary to deuterate the protein to improve its relaxation properties. Deuteration of the protein results in all ^1H being replaced by ^2H that has a much slower T_2 relaxation, which results in a significant decrease in line broadening. On the other hand the regular CBCA pair does not work and one has to use the less sensitive HNNCACB and HNN(CO)CACB (called out-and-back) experiments[30]. Nevertheless this will be able to give correlated chemical shifts that could not be detected before deuteration, due to the improved relaxation properties.
In principle it is good to use the highest available field (750 MHz or higher) to improve the resolution of the spectra and to also take advantage of TROSY techniques[31,32]. The TROSY technique is based on that peaks appear as multiplets due to J-coupling, when decoupling is not applied. Crucially the different multiplet components have different widths. This is due to constructive or destructive

interaction between different relaxation mechanisms. Typically relaxation for large proteins at high magnetic field strengths, the transverse T_2 relaxation is dominated by the dipole-dipole (DD) mechanism and the chemical shift anisotropy (CSA) mechanism. As the relaxation mechanisms are generally correlated but contribute to the overall relaxation rate of a given component with different signs, the multiplet components relax with very different overall rates. The TROSY experiment is designed to select the component for which the different relaxation mechanisms have almost cancelled, leading to a single, sharp peak in the spectrum (Figure A.1.2.1.1).

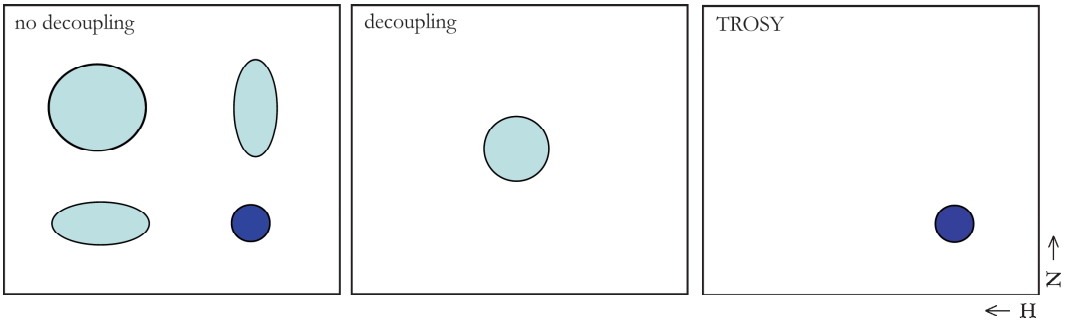

Figure A.1.2.1.1. When no decoupling is applied in a ^{1}H-^{15}N HSQC for a large system at high fields the different relaxation properties can be seen on the multiplets, left panel. For one of the components the relaxation mechanisms have almost cancelled, blue circle left panel, leading to a sharp peak. The other three components experience different relaxation rates due to the dipole-dipole and chemical shift anisotropy mechanism effects. The middle panel depicts the same experiment with decoupling on the ^{15}N nuclei during acquisition and decoupling of ^{1}H during ^{15}N evolution. The TROSY experiment selects for the component, where the CSA and dipole-dipole relaxation mechanisms cancel out (blue circle right panel).

This significantly increases both spectral resolution and sensitivity. But, it generally requires high magnetic fields (> 750 MHz) to achieve the necessary balance between the CSA and DD relaxation mechanisms; CSAs scale with field strength, while dipole-dipole couplings are field-independent.
Using perdeuteration and TROSY techniques has proven to be a very powerful way of acquiring high quality data for large proteins[30].
The basis of the CBCA pair 3D spectra is to correlate the chemical shifts of all the residues and link them to each other. The CBCANH experiment correlates the NH group with its Cα, Cβ chemical shifts of its own residue (strongly) and of the preceding residue (weakly). The CBCACONH only correlates the NH group to the preceding Cα, Cβ (Figure A.1.2.1.2).

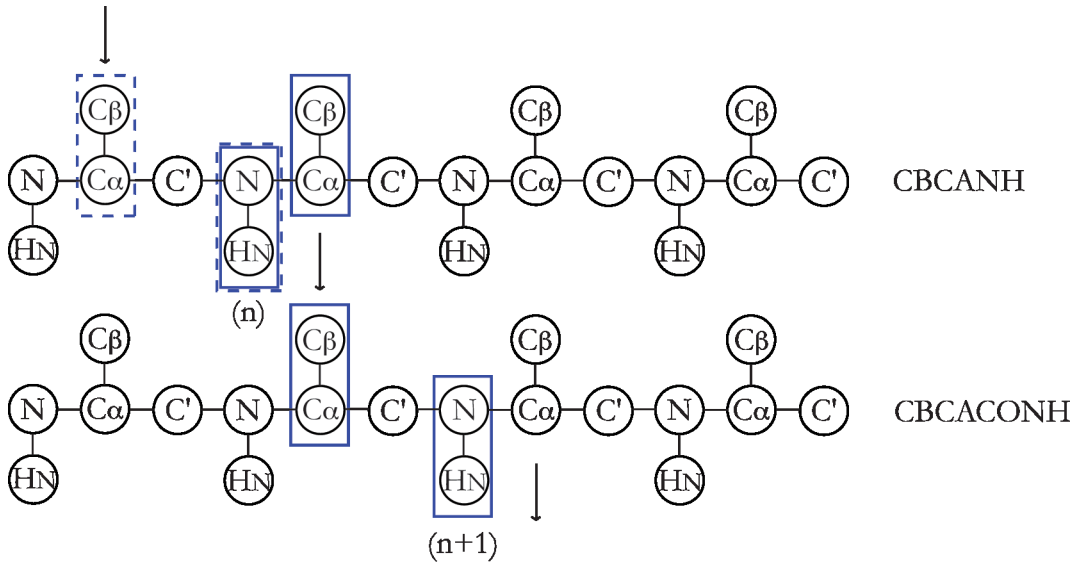

Figure A.1.2.1.2. Depicts the peptide chain and the resonances visible for each of the CBCA pair experiments, CBCANH top and CBCACONH bottom. Marked in blue boxes are the visible resonances that are correlated to each other, Cα, Cβ to N, HN. Resonances in dashed boxes are visible at lower intensity. Two correlations are obtained from the CBCANH, n and n-1, and in the CBCACONH, n+1. After a Cα, Cβ pair has been determined in CBCANH the CBCACONH experiment can be used to correlate these resonances to the next neighbour residue in the chain. The resonances for the next residue is then displayed in the CBCANH and are matched to the subsequent residue. This procedure is called "walking through the chain".

The most common way of visualising 3D spectra is with so-called Strips. If one starts from the easier 2D ^{1}H-^{15}N HSQC spectra, every peak correlates to a residue defined by the ^{1}H, ^{15}N chemical shifts. The coordinates (chemical shifts) in the 3D spectral cube will yield a plane, usually with ^{1}H chosen as x-axis, with the Cα, Cβ, chemical shift, shown in the ^{13}C dimension, y-axis. Each peak in the 2D ^{1}H-^{15}N HSQC will hence yield a 2D strip with its H^{N} (^{1}H), Cα and Cβ chemical shifts (^{13}C). This is better understood in Figure A.1.2.1.3.

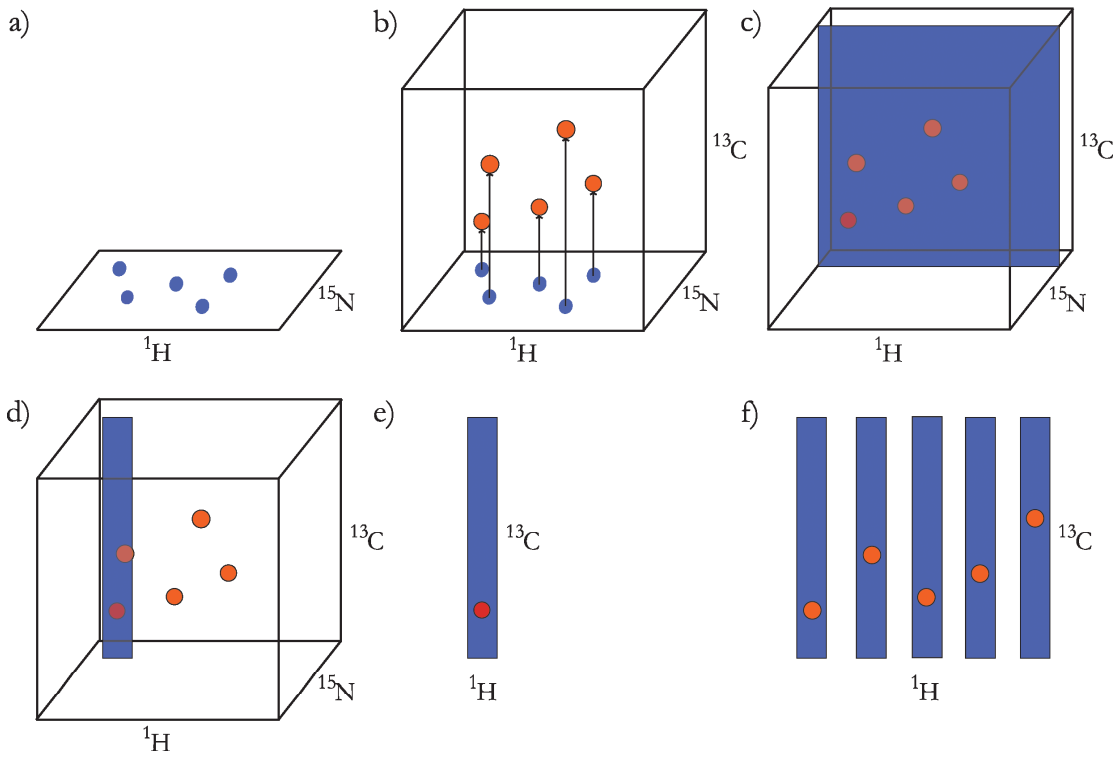

Figure A.1.2.1.3. Shown is how a strip plot is selected from a 3D experiment. The ^{1}H-^{15}N HSQC serves as a projection of the ^{1}H, ^{15}N dimension of the 3D experiment, a). The resolution of the correlation in the 3D experiment in comparison to the 2D is greatly enhanced, b), and by choosing a plane in the ^{15}N dimension one can easily obtain the ^{1}H, ^{13}C correlation of respective residue in the ^{1}H-^{15}N HSQC, c). The plane is then cut to only obtain the correlations of interest, d). When all the strips have been collected they can be put next to each other for comparison, e) and f).

From the reasoning from Figure A.1.2.1.2 all the strips are connected to each other, Figure A.1.2.1.4. Remember that proline due to its build up does not show up in the 2D ^{1}H-^{15}N HSQC spectra or in the NH plane of the CBCA pair. But, the Cα, Cβ can be detected from the preceding residue.

The Cα, Cβ chemical shifts adopt values characteristic of the amino acid type. Alanine, Serine, Threonine and Glycine are very easy to determine due to their Cβ chemical shifts being very different of those of the other amino acids (In the case of Glycine there is no Cβ). Valine, Isoleucine are also likely to stand out by the fact that they have lower than normal Cα chemical shifts. Once a chain of NH groups with their corresponding Cα and Cβ chemical shifts has been built, the identification of some of the amino acid types makes it possible to match this string to the sequence. If some of the residues in the chemical shift matching string can be readily identified, i.e. xxxGxxxTx the two readily assigned residues can be matched to the primary sequence and sequence specific assignment can be made. Note that, without the primary sequence the sequence specific assignments can't be made!

In larger systems the signal-to-noise or the spectral overlap can be to poor or to great respectively. In these cases the HNCA/HNCOCA pair and the HNCO/HNCACO pair experiments can be used. They both give reduced spectra and are more sensitive since

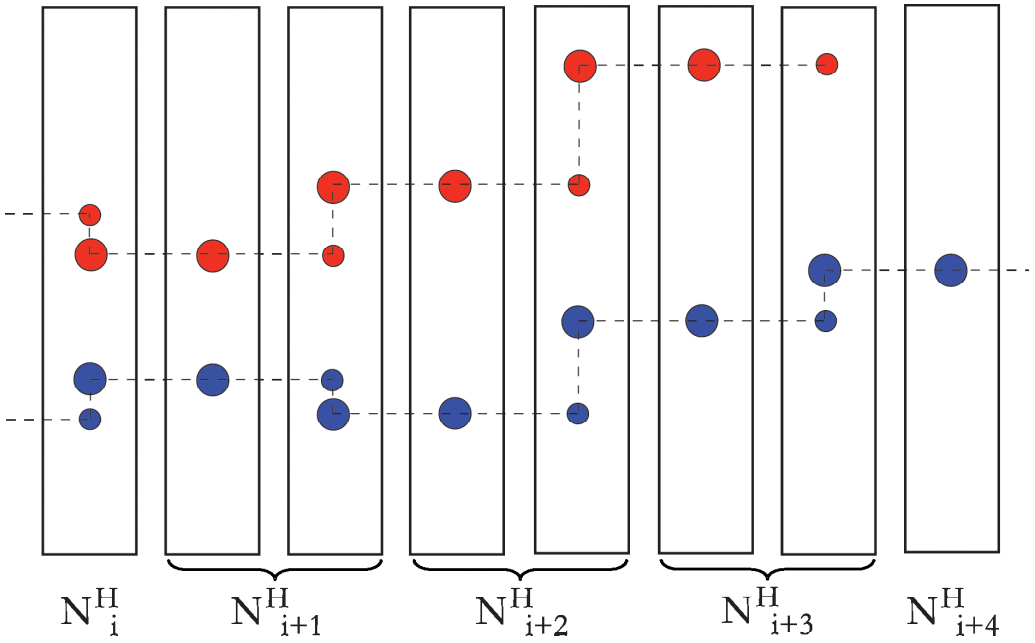

Figure A.1.2.1.4. Chemical shift matched strips of the CBCA pair experiments. The Cα, Cβ chemical shifts of N^H_i are shown in a CBCANH strip in the very left panel. The Cα, Cβ chemical shifts are then correlated to the N^H_{i+1} strip from the CBCACONH. Strips are then matched to complete the "walk through the chain".

the observed magnetisation only comes from one correlation and not two. Though the HNCO experiment is the most sensitive 3D experiment in protein NMR, the dispersion of the C' nuclei is very small, 174 – 177 ppm.

A.1.2.2 Sidechain assignments

Various methods and spectra are available for sidechain assignments, which method is the best is debatable and comes down to personal taste. The classical way is to record the HCCH-COSY and HCCH-TOCSY experiments[11]. The HCCH-TOCSY will at any one carbon position show in one dimension the chemical shift of the hydrogen which is attached to the carbon and in another the other hydrogens belonging to that sidechain. The dispersion of the carbon chemical shifts ranges between 10 - 80 ppm, and the irradiation of the carbon channel is usually set at around 45 ppm. This has the consequence that the Cα, Cβ carbons have higher signal-to-noise and sometimes the methyl carbons of Leucine can have poor signal intensities. The TOCSY transfer of magnetisation in the carbon chain is not perfect and the transfer might not be enough to see the full sidechain from the Cα or Cβ chemical shift. The HCCH-COSY is similar, except that instead of transferring the magnetisation through the whole sidechain, it only transfers magnetisation to the neighbouring carbon atoms. The general principle behind using the HCCH-TOCSY and HCCH-COSY spectrum is that from the already known Cα, Cβ chemical shifts correlate them to the carbon chemical shifts in the HCCH-COSY, HCCH-TOCSY. The HCCH-COSY, HCCH-TOCSY strips will then contain the hydrogen chemical shifts of the sidechain. By navigating

via these new hydrogen chemical shifts, both ^{1}H dimensions is set to a hydrogen shift i.e. Hγ the carbon chemical shift correlated to that hydrogen will be on the diagonal in the spectrum, the rest of the carbon chemical shifts can be identified. This is illustrated in Figure A.1.2.2.1 for a Valine residue.

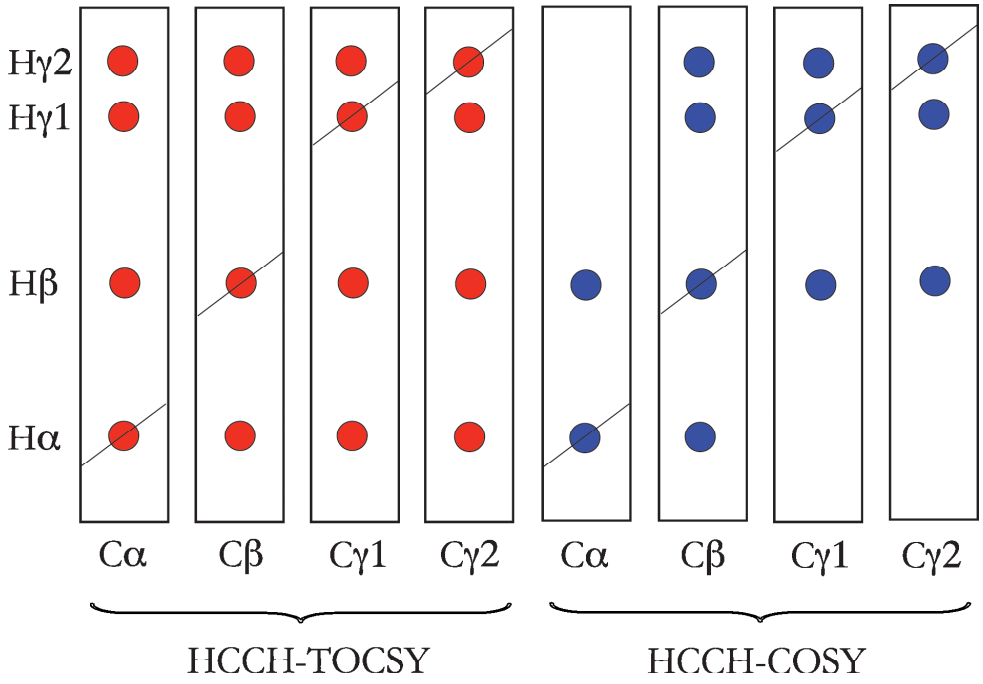

Figure A.1.2.2.1. Visible correlating cross-peaks in a Valine residue depending on the different transfer of magnetization used in a HCCH-TOCSY and HCCH-COSY, left and right respectively. Strips are set to the different auto-correlations of Cα, Cβ, Cγ1, Cγ2 of Valine, left to right respectively, for both the HCCH-TOCSY and HCCH-COSY experiments[11].

When using perdeuterated proteins to obtain Cα, Cβ chemical shifts these need to be corrected for in the HCCH-TOCSY and HCCH-COSY experiments. The deuterium nuclei attached to the carbons will perturb the chemical shifts of the Cα, Cβ nuclei in residue specific manner, Table A.1.2.2.1, and its effect has been reported by Gardner and co-workers[33].

Table A.1.2.2.1. Chemical shift difference between the ^{15}N, ^{13}C, ^{2}H and the ^{15}N, ^{13}C, ^{1}H domain of PLCC SH_2[a33].

Amino acid	^{15}N	$^{13}C\alpha$	$^{13}C\beta$
Gly	-0.23 ± 0.10	-0.50 ± 0.02	Na
Ile	-0.16 ± 0.12	-0.45 ± 0.04	-0.91 ± 0.07
Thr	-0.15	-0.40	-0.57
Val	-0.23 ± 0.07	-0.49 ± 0.05	-0.77 ± 0.02
Asn	-0.28	-0.34 ± 0.01	-0.61 ± 0.01
Arg	-0.17 ± 0.14	-0.43 ± 0.02	-0.70 ± 0.02
Cys	-0.49	-0.46	-0.65
Gln	-0.19 ± 0.05	-0.43 ± 0.04	-0.82 ± 0.05
Glu	-0.22 ± 0.07	-0.44 ± 0.09	-0.83 ± 0.06
His	-0.18 ± 0.05	-0.44 ± 0.08	-0.72 ± 0.05
Leu	-0.28 ± 0.06	-0.45 ± 0.05	-1.08 ± 0.04
Lys	-0.22 ± 0.06	-0.40 ± 0.04	-1.04 ± 0.06
Met	-0.29 ± 0.05	-0.46 ± 0.10	-0.94 ± 0.03
Phe	-0.24 ± 0.08	-0.37 ± 0.03	-0.90 ± 0.01
Pro	nd	-0.45 ± 0.01	-0.85 ± 0.05
Ser	-0.20 ± 0.10	-0.43 ± 0.09	-0.72 ± 0.06
Trp	-0.15	-0.34	-0.69
Tyr	-0.27 ± 0.11	-0.43 ± 0.05	-0.86 ± 0.04
Ala	-0.22 ± 0.05	-0.42 ± 0.04	-0.93 ± 0.05

Sidechain nitrogen chemical shifts, belonging to Lysine, Arginine, Glutamine, Asparagine, Tryptophan and Histidine, cannot be assigned with previous mentioned experiments. Nitrogen chemical shifts of Glutamine and Asparagine resonate around 112 ppm and both attached hydrogen chemical shits resonate between 6.8 - 7.8 ppm and can easily be observed in the ^{1}H-^{15}N HSQC. The Tryptophan nitrogen chemical shifts resonate at around 129 ppm and the hydrogen chemical shift between 11 – 9.8 ppm. Lysine Nζ and Arginine Nη chemical shifts resonate at around 30 and 85 ppm, and with their hydrogen chemical shifts at 7.8 and 6.2 – 7.0 ppm, respectively. The nitrogen chemical shifts are outside of the normal spectral window, but are observed folded into the visible spectral window. The Histidine nitrogen chemical shifts resonate between 180 – 200 ppm, and the hydrogen chemical shifts between 7.0 – 13 ppm. Therefore specially constructed experiments needs to be implemented to detect them.

The connectivity between the sidechain nitrogens and the amide nitrogen can be seen in the ^{15}N-NOE-HSQC[11] by matching strip plots from the N-H^N correlations in the ^{1}H-^{15}N HSQC. If multiple matching NOE peaks can be seen in the amide and sidechain nitrogen strip plots, they correlate each other and the resonances can be readily assigned. Strip plots for two Tryptophan residues with their Nε strips are illustrated in Figure A.1.2.2.2 together with ranges plotted out in a ^{1}H-^{15}N HSQC where the sidechain nitrogens commonly resonate.

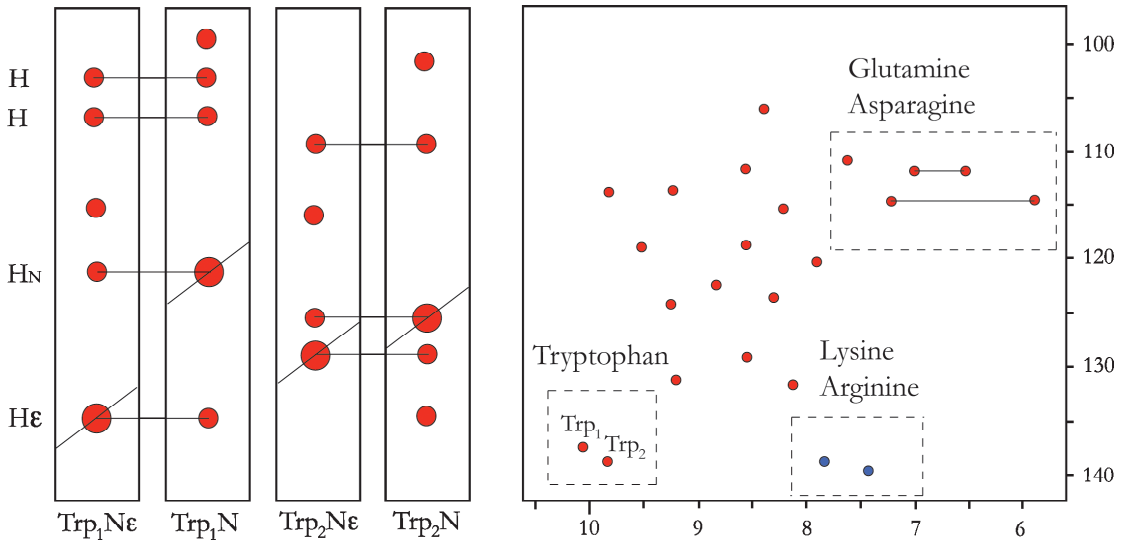

Figure A.1.2.2.2. NOE peaks are matched in the N, Nε strip plots (left panels) and if correlating peaks can be seen, the strip scan be matched and assigned. Right panel depicts the areas where the sidechain nitrogens of Trp, Lys, Arg, Gln and Asn normally resonate. Lys $^{15}N_{\zeta}$ chemical shifts resonate at ~30 ppm but are doubly folded into the spectrum at ~135 ppm (blue). Arg $^{15}N_{\eta}$ chemical shifts resonate at ~85 ppm, and are folded into the spectral window at ~140 ppm (blue).

Other common experiments for sidechain assignments are the HBHACONH[11], that correlates the Hα, Hβ chemical shifts to the N-H^N resonances, the CCCONH[11] that correlates the carbon chemicals to the N-H^N, the HACACONH[11] can give the Hα chemical shifts of bigger proteins that are not deuterated. The advantage with these three experiments is that they correlate hydrogen and carbon values directly from the N-H^N chemical shifts, and hence the deuterium chemical shift correction step is not needed.

Random coil chemical shifts and chemical shifts of well-folded proteins have been gathered creating a database table with expected chemical shift values and their ranges for amino acids in folded and unfolded states. One of these can be found at the BMRB databank for NMR at http://www.bmrb.wisc.edu/devise/histogram.html. These reported values will greatly aid in the assignments of the chemical assignments and should be used to validate the complete assignment.

A.1.3 Secondary structure from chemical shifts

NMR can provide structural data by assessment of the assigned chemical shifts from a macromolecule. As described before, the chemical shifts of amino acids are not only highly sensible to the amino acid type, but also to the local conformation of that amino acid. Secondary structures give distinct perturbations of the chemical shifts. Random coil structure, dipeptides or tripeptides of all natural amino acids, chemical shifts have been measured and analysed against chemical shifts of proteins with solved structure. This comparison between measured chemical shifts from a protein without a known structure to the random coil values can give important early structural data without too much effort. This relationship between secondary structure

and chemical shifts was first described in the late 1960s by Jardetzky and co-workers[34] and has been given a renaissance by Wishart and co-workerks in 1992[35,36]. Solving the secondary structures is the very first step into getting the full tertiary structure. Distinct motifs can be recognised and later these smaller structural components will be the starting point of the arrangement of these within the overall tertiary structure. Two different methods of doing secondary structure predictions are described further in this chapter.

A.1.3.1 Direct interpretation of chemical shifts

The most straightforward method is to use the Cα, Cβ chemical shifts of all assigned amino acids in a protein sequence[35-37]. The chemical shifts are directly compared to the random coil chemical shifts and the difference can then be plotted to determine the similarity to any of the three secondary structure elements. An alpha helical secondary structure gives low field Cα chemical shifts and up field Cβ chemical shifts respectively. Beta strands on the other hand perturbs the chemical shifts up field for the Cα and low field for the Cβ. There is no direct assessment of the $_3$10 helices, but these can be readily detected in the ^{15}N-NOE HSQC experiment. A minimum stretch of three amino acids in the primary sequence should show the same behaviour to be considered as a secondary structure element. Longer stretches are more reliable than short ones and local mismatches can be acceptable, i.e. a six stretch with the first two amino acids showing a clear beta strand dependence followed by one amino acid with no beta strand similarity followed by three amino acids with a clear beta strand similarity. This is most likely a six amino acid long beta strand. Examples of beta strand and alpha helical chemical shift predictions are illustrated in Figure A.1.3.1.1.

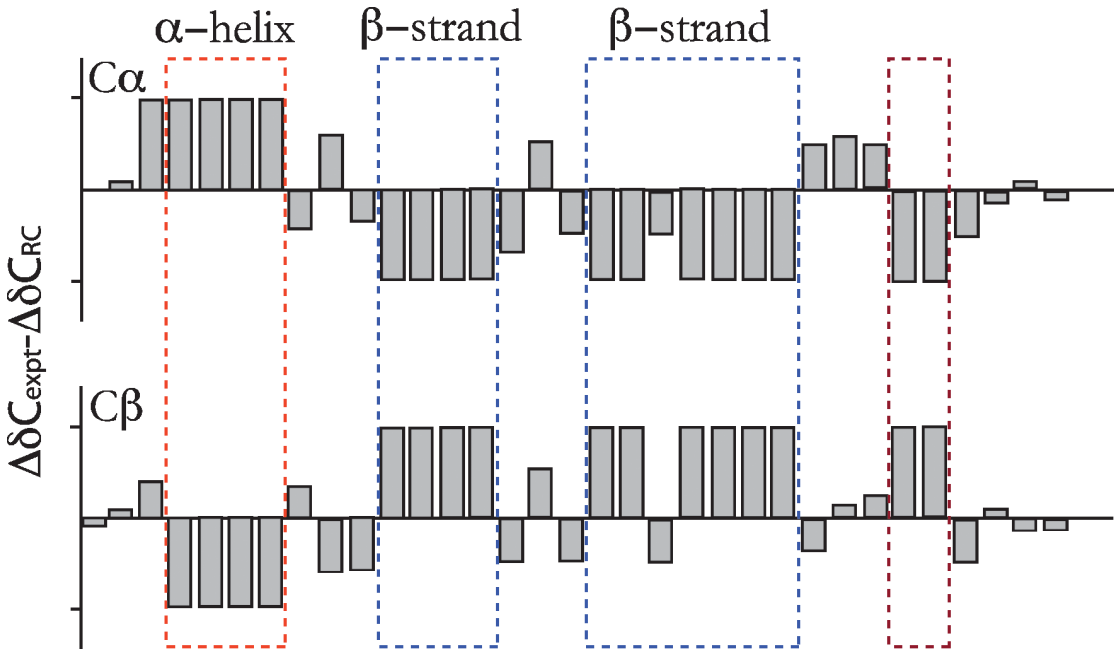

Figure A.1.3.1.1. Cα, Cβ chemical shift comparison. α-helical propensity can be seen for the amino acids in the red box, beta strand propensity for amino acids in blue box. Right blue box contains one set of Cα, Cβ shifts that does not match a beta strand, but are considered an outlier and belonging to the

beta strand. In burgundy a clear beta strand propensity for a too short stretch, i.e. not considered a beta strand.

A.1.3.2 Database analysis of chemical shifts

Where the direct analysis of the chemical shifts for secondary structure propensities is a rough tool other methods have been developed. The software TALOS, Cornilescu *et al.*[38], calculates protein backbone torsion angles by inference from measured Cα, Cβ chemical shifts and prior information available at BMRB and PDB databases. TALOS is a database system for empirical prediction of phi and psi backbone torsion angles using a combination of five (CO is optional) kinds (Hα, Cα, Cβ, CO, N) of chemical shift assignments for a given protein sequence. The approach is an extension of the well known observation that many kinds of secondary chemical shifts (difference between measured chemical shifts and their corresponding random coil values) are highly correlated with aspects of protein secondary structure.

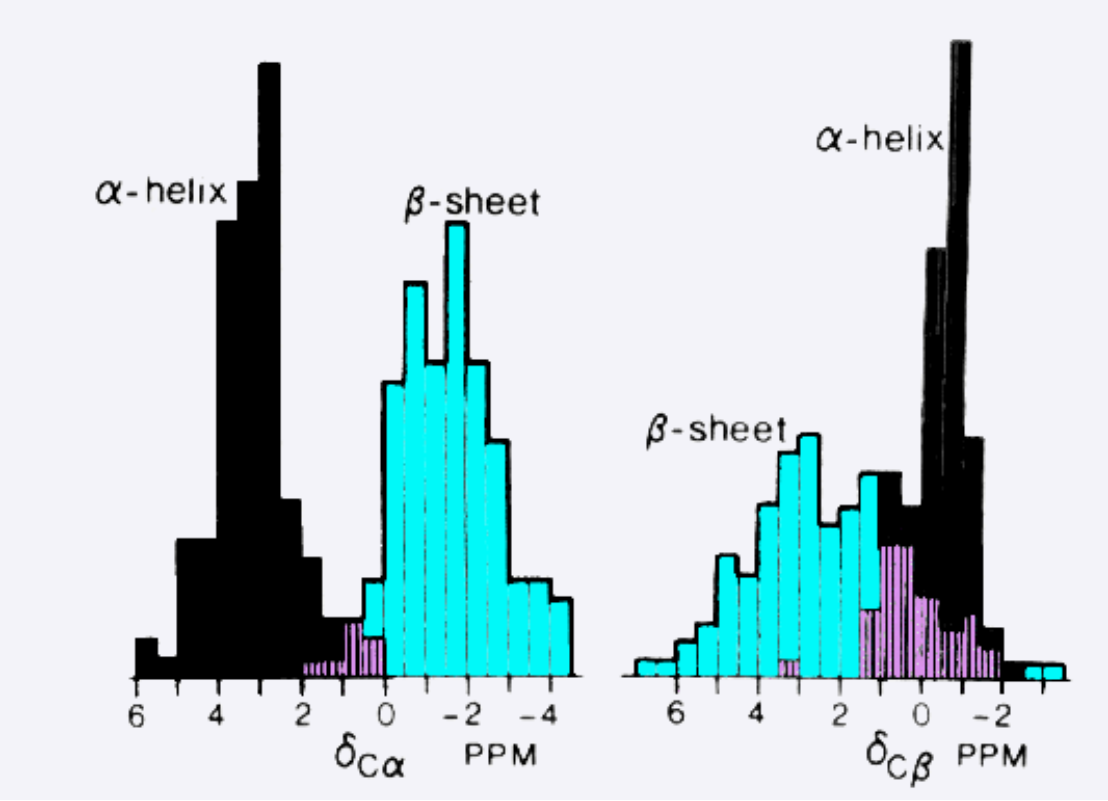

Figure A.1.3.2.1. Cα and Cβ secondary shifts distribution and their propensity for secondary structure type. Figure from Cornilescu *et al.*[38].

TALOS use the secondary shift and sequence information in order to make quantitative predictions for the protein. The secondary shifts of a given amino acid with the secondary shifts of the preceding and next amino acids are used; data from three consecutive amino acids are used simultaneously to make predictions for the central amino acid in a triplet. The triplet is matched to a protein with known structure and if the secondary shifts are also well matched, and then the phi and psi angles in

the known structure will be useful predictors for the angles in the target. The similarity is measured with a score based on the weighted sum of squares differences between the shifts in the target protein and the database entries, so that lower scores indicated high similarity. In order to take advantage of the correlations between residue type and secondary structure, the score also includes a small, qualitative residue type term, which biases the matching towards roughly similar sequences. TALOS searches a database for the 10 best matches to a given triplet in the target protein. If these 10 matches indicate consistent values for phi and psi, then their averages and standard deviations are used as a prediction. However, if the 10 best matches have mutually inconsistent values of phi and psi, the matches are declared ambiguous, and no prediction is made for the central residue. In the TALOS approach, an initial classification of good versus ambiguous is performed automatically, and the classifications are then adjusted interactively through a graphical interface, which is part of the TALOS system.

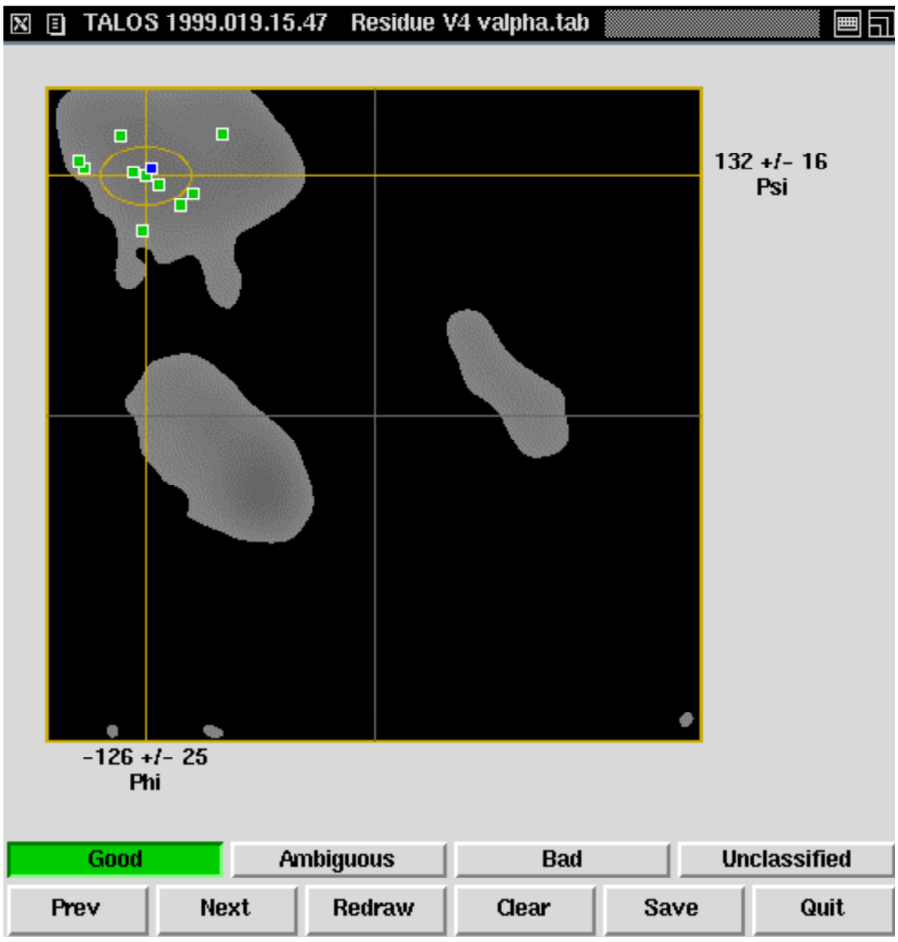

Figure A.1.3.2.2 Ramachandran plot with predicted values of backbone angles for a specific amino acid. Green squares represent the ten matched secondary shifts to ten different structures in the TALOS database[38]. All matches are within the β-strand phi- and psi-angles of the Ramachandran plot and therefore have a strong prediction for β-strand propensity in the unknown structure. Figure from Cornilescu *et al.*[38].

The TALOS database, while small, was constructed using the most well defined parts of high resolution (2.2 Å or better) X-ray crystal structures to define the phi and psi angles. It originally included data from 21 proteins, representing around 3,000 triplets. The current database includes data from 186 proteins, representing over 24,000 triplets.
The accuracy of TALOS predictions have been cross-validated by removing one of the structures from the database and then predicting its phi and psi angles from the secondary shift data. The result were that TALOS makes predictions for 72 % of the residues on average, with a range from 55 – 80 %, over all 186 protein structures, about 1.8 % of the predictions were incorrect relative to the corresponding protein structure.
TALOS backbone angle restraints are easily attained and create a vast amount of data for structure calculation of NMR protein structures. The TALOS angle restraints have been incorporated into all common structure calculation programs and have been a great tool for structure determination not only for systems with limited distance restraints data.

A.1.4 Chemical shift mapping

Protein-protein interactions are operative at almost every level of cellular function, in the structure of sub-cellular organelles, the transport machinery across the various biological membranes, packaging of chromatin, the network of sub-membrane filaments, muscle contraction, signal transduction, and regulation of gene expression to name a few. Aberrant protein-protein interactions are implicated in a number of neurological disorders such as Creutzfeld-Jacob and Alzheimer's disease.
Because of their importance in development and disease, these systems have been the object of intense research for many years. It has emerged from these studies that nature has employed in many instances a strategy of mixing and matching of domains that specify particular classes of protein-protein interactions. In the current post-genomic era these domains and classes can be readily identified and vast amount of systems can be interpreted in a rapid pace. To fully understand all these processes a large variety of methods and analysis can be used for elucidation. A review by Fields and co-workers[39] shows a wide range of biochemical methods at hand, and numerous structural biology incentives like RIKEN[40], Montreal-Kingston Bacterial Structural Genomics[41] and the CESG[42], have helped to give complimentary high-resolution data.

The NMR chemical shift is extremely sensitive to any changes in its vicinity; some of the effects that can alter it has been described in chapter A.1.2.1 NMR Theory. There are many ways of detecting protein-protein or protein-ligand interactions by NMR; Pascal *et al.*[43] have compiled some of the methods. A common way is to record 2D ^{1}H-^{15}N HSQC or ^{1}H-^{13}C HSQC experiments in absence and presence of the ligand. The chemical shift perturbations display the interaction and if a 3D structure is available the interaction site can be mapped out on the protein surface. The chemical shift perturbations can also be used to determine the K_D value of the interaction[43]. With an uniformly ^{15}N labelled protein, interaction studies can be performed with

proteins < 20 kDa[44], with ^{15}N, ^{1}H labelled proteins this size limit can be extended to < 80 kDa, or in some extremely favourable cases even larger protein complexes, with the help of earlier mentioned TROSY-type experiments[31,45,46]. Flexible regions of interest of very large proteins like the 20S proteosome (670 kDa)[47] can also be studied if the dynamics is within NMR time scale[14,48].
HSQC chemical shift mapping requires relatively high protein concentrations ranging between 50 – 500 μM, but other NMR techniques have been developed to overcome this. STD-NMR[49], WaterLogsy[50,51], NOE pumping[52] and transfer NOE[53], all uses unlabelled protein and ligand and at concentrations of 10 – 100 μM, but does not always give high-resolution information. Nevertheless, NMR is able to perform as a robust method that gives the strength and localization of the interaction. This with little effort and a small amount of time.

A.2 NMR relaxation

In the recent decade X-ray diffraction and multidimensional NMR have determined a remarkable number of three-dimensional structures of proteins and nucleic acids. While this effort has provided a wealth of insight into protein architecture of reactive sites and biomolecular interfaces, it has also become abundantly clear that knowledge of the static structure alone is insufficient to explain how biomolecules function. Perhaps as a consequence, designing therapeuticals on the basis of static biomolecular structures has been much less successful than expected. Many lines of evidence point to the existence of large scale dynamics on the milli- to microsecond time scale that are essential to biomolecular function. Even events as basic as the binding of oxygen to hemoglobin are only possible when the protein transiently opens. Entire enzyme domains must move to sequester and release substrates. Hormones deliver their commands to the cellular nucleus by a signal transduction cascade of protein–protein interactions, in which the proteins dynamically change their conformations between ‘‘on’’ and ‘‘off’’ states, R_{ex}[117]. The amplitudes of the above motions can almost be as large as the enzymes themselves. It is of interest to study the time scale of these motions/conformational changes, and it would be the dynamicist’s dream to actually describe these motions in terms of models. Much faster dynamics, on the nano- to picosecond time scale, are equally interesting. Many experimental results show that these motions are present, from temperature factors that accommodate vibrations in atom positions in X-ray structures, from molecular dynamics simulations, and of course, from NMR relaxation. The motions can be as large as several tenths of an Ångstrom, and can be associated with retention of entropy in the folded state and with the protein’s

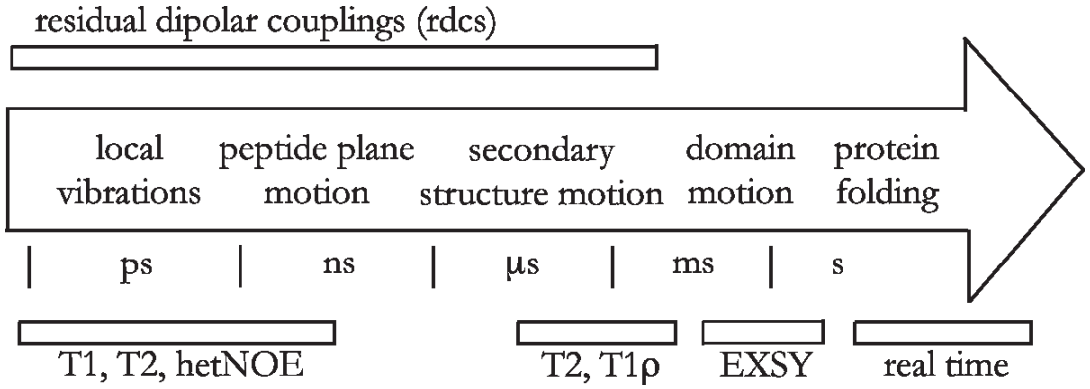

Figure A.1.2.1. Time-scales of high resolution NMR. The different dynamic processes with their time scale, arrow, and the NMR properties they affect, boxes.

management of thermal energy in general. The presence of dynamics may appear at odds with the protein's local functions. Proteins often carry out chemical reactions, which require orbital overlap between the reacting species. Orbital overlap is a very steep function of inter-atomic distance, and (random) dynamical events of several tenths of an Ångstrom in amplitude seem large compared to the precision required. Is it possible that the fast motion of individual protein atoms allows active-site residues in enzymes to make transient contact with the substrate to affect the enzyme's chemistry, thus circumventing the need for more precise design/evolution of the active-site architecture? Or can it generally be stated that flexibility in the active site may guide the chemical dynamics of transition states, as was suggested for Bacteriorhodopsin[20]. Alternatively, are the motions of active-site atoms correlated and are essential chemical groups at the static nodes of more global normal modes[21]. Local geometry is maintained if atoms move in a correlated way, but not when uncorrelated motion pertains. In locations with correlated motion, chemistry and function could be understood on the basis of a static structure description, in locations with uncorrelated motion they could not.

To understand how these effects can be measured one has to understand the different NMR relaxation processes. The term relaxation in NMR describes the rate by which nuclear magnetization prepared in a non-equilibrium state return to the equilibrium distribution. The relaxation rate describes how fast spins reorient themselves to ground state. Different physical processes are responsible for the relaxation of the components of the nuclear spin magnetization vector, parallel and perpendicular to the external magnetic field, B_0, which is conventionally oriented along the z-axis. The principal relaxation processes are termed longitudinal or spin-lattice, T_1, and transverse or spin-spin, T_2, for an isotropic system. Relaxation due to cross-relaxation mechanisms is named the *nuclear-Overhauser-effect* (*NOE*). The rotational relaxation time, $T_{1\rho}$, is the same as T_2 unless there is chemical exchange or anisotropy in the system. The parameter T_2^* is the time constant of dephasing and is caused by a combination of relaxation and magnetic field inhomogeneity.
The theoretical treatment of NMR relaxation phenomena may be carried out on many levels ranging from a fully classical to fully quantum mechanical treatment. The fully

classical treatment correctly predicts the general behaviour of spin-lattice or longitudinal relaxation (R_1), but is however unable to generate the correct analytical expression for the longitudinal relaxation time, T_1. A semi-classical treatment, in which the relaxing spin is considered quantum mechanically, but the lattice or thermal bath is treated classically, gives the correct expression for the relaxation time. If the Redfield density matrix formalism is used, this approach also leads naturally to a description of spin–spin (R_2) relaxation. This semi-classical approach, however, fails because it predicts that as the system relaxes, all states become equally populated, equal to an assumption of infinite temperature. A fully quantum mechanical treatment of the spins and the lattice or thermal bath results in an approach to the correct equilibrium situation given by the Boltzmann distribution. The different treatments of the relaxation phenomena therefore make the theory very complex and would also expand this chapter beyond its point. Immense work, over a long time, describing these phenomena, theoretically and empirically, has already been published to great length by Fischer *et al.* 1998[22], Palmer A. G. 2001[23], Cavanagh *et al.* 2007[11] to name a few.

A.2.1 T_1-relaxation

Consider the proton nuclei in a sample of water. In the absence of an external magnetic field, the spin polarizations are uniformly distributed, pointing in all possible directions in space. The total magnetic moment of the sample is very close to zero, since approximately the same number of spins point towards a given direction as against it.
If a magnetic field is suddenly turned on, all protons spins begin executing Larmor precession around the field. This precessional motion is essentially invisible. It does nothing to change the total magnetic moment of the sample. An isotropic distribution of spin polarizations makes no contribution to the magnetism of the material. As the external magnetic field is turned on the nuclei will start to grow to their biased spin polarizations (Larmor frequency) by a build up curve. Suppose that t_{on} is defined as the moment when the external magnetic field is applied, and the direction is defined to be the z-axis. The build up of longitudinal magnetization has the form:

$$M_z^{nuc}(t) = M_{eq}^{nuc}\left(1 - \exp\left\{-\left(t - t_{on}\right)/T_1\right\}\right) \qquad [A.2.1.1]$$

for times $t \geq t_{on}$:

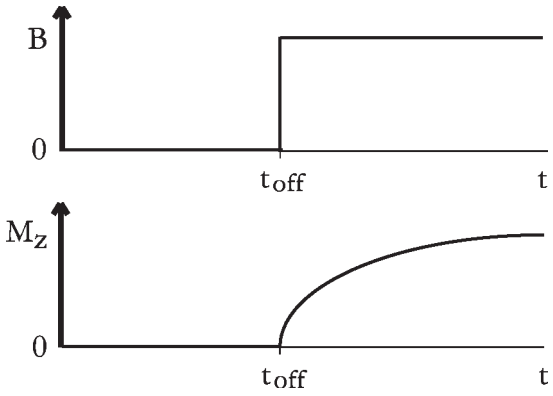

Figure A.2.1.1. The buildup of longitudinal spin magnetization. Figure from Spin dynamics: Basics of Nuclear Magnetic Resonance[102].

The exponential time constant for the process T_1 is known as either the spin-lattice relaxation time constant or the longitudinal relaxation time constant. The use of the term 'lattice' derives from the early days of NMR, when theoretical efforts concentrated on the treatment of NMR in solids, and thermal equilibration was explained in terms of the interactions between the nuclear spins and the crystal lattice. The term is still used in NMR for liquids and gases, with the term 'longitudinal' indicating that the magnetization build up occurs in the direction of the applied magnetic field.

The term 'relaxation' is widely used in the physical sciences to indicate the re-establishment of thermal equilibrium after some perturbation is applied. In the case discussed above, thermal equilibrium is first established in the absence of an external magnetic field, so that all spin orientations are equally likely. When a magnetic field is applied, this situation no longer corresponds to equilibrium, and the system 'relaxes' to the new equilibrium state, in which the spin polarizations are distributed anisotropically. If the magnetic field is suddenly turned off at a later time, t_{off}, where $t_{on} - t_{off} \gg T_1$, the nuclear spin magnetization relaxes back to zero again, following[102]:

$$M_z^{nuc}(t) = M_{eq}^{nuc} \exp\left\{-\left(t - t_{off}\right)/T_1\right\} \qquad [A.2.1.2]$$

For times $t \geq t_{off}$:

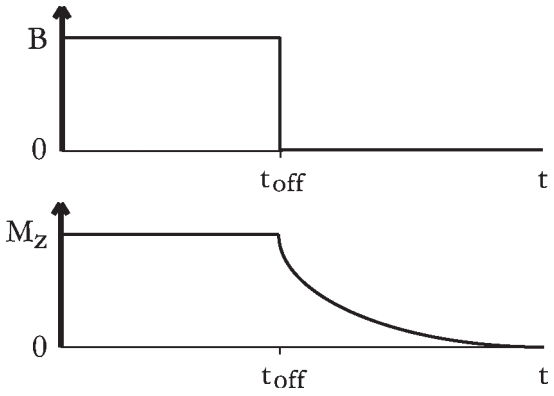

Figure A.2.1.2. The decay of the longitudinal spin magnetization. Figure from Spin dynamics: Basics of Nuclear Magnetic Resonance[102].

or shown in a more complex manner of spectral density, the energy of the signal distributed by the frequency, versus the correlation time, τ_c, of a molecule[118]:

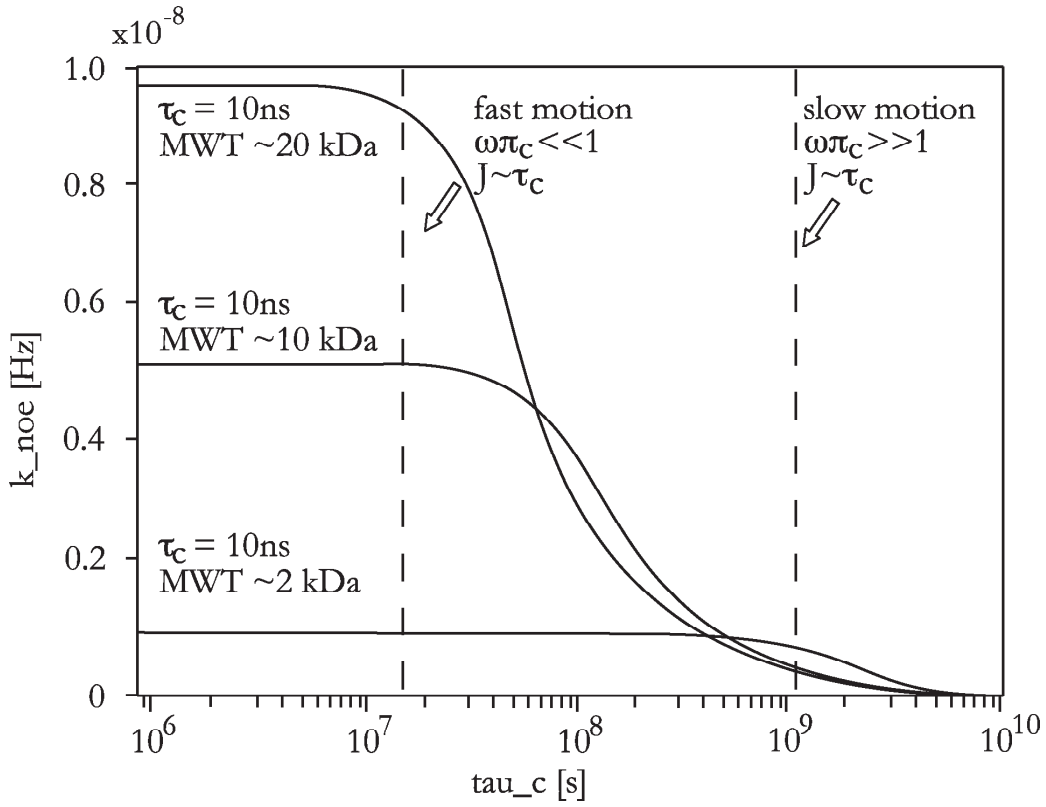

Figure A.2.1.3. The spectral density or the efficiency of relaxation for different Larmor frequencies and rotational correlation times τ_c. Figure from Grzesiek S. 2005[118].

The relaxation time constant T_1 depends on the nucleus and the sample, including parameters such as temperature and viscosity, if the sample is a liquid. Typically the value of T_1 is in the range of ms to s, although T_1 may be as long as days or even years in exceptional cases.

Measuring the T_1-relaxation
For spin-lattice RS(S_z) measurements, an inversion-recovery building block is applied during the relaxation building block. Thus, after an initial refocused INEPT pulse train, the in-phase S magnetization is converted to S_z magnetization by applying a 90^0 S pulse. During the T period, proton saturation is needed to remove other relaxation mechanisms. After the following T period, another 90^0 S pulse creates transverse S magnetization, which is allowed to evolve during the conventional t_1 period and converted to the desired in-phase ^{1}H magnetization by a refocused retro-INEPT pulse train.
The signal intensities are then measured for different T periods. Fitting of the signal intensities gives the T_1-relaxation for the measured system.

A.2.2 T_2-relaxation

The longitudinal nuclear spin magnetization is almost undetectable[12,102]. It is about four orders of magnitude less than the typical diamagnetism of the sample, associated with the electrons. Experimental study of the longitudinal nuclear magnetization is impractical.
Instead of measuring the nuclear spin magnetization *along* the field, the magnetization *perpendicular* to the field is measured. The net magnetization perpendicular to the field is achieved by applying a radiofrequency pulse, causing the spins to rotate $\pi/2$ radians around the x-axis, of appropriate frequency and duration. Only the consequences of this will be discussed.
If a spin polarization is initially along the z-axis, and is rotated by $\pi/2$ about the x-axis, the result is a spin polarization along the –y-axis.

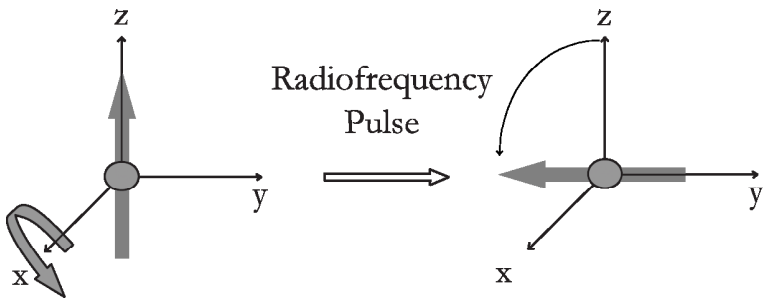

Figure A.2.2.1 Rotation of a spin around the x-axis. Figure from Spin dynamics: Basics of Nuclear Magnetic Resonance[102].

Since the pulse rotates the polarization of every single spin in the sample by the same angle, the pulse also rotates the entire nuclear magnetization distribution of the sample. The net spin polarization along the z-axis is therefore transferred into a net spin polarization along the –y-axis.
This net magnetization moment perpendicular to the magnetic field is called the

transverse magnetization.
When the pulse is turned off the spins continue their precessional motion. The state immediately after the pulse corresponds to a net polarization along the –y-axis, perpendicular to the main field. Since every single spin precesses, the bulk moment precesses too. The macroscopic nuclear magnetization rotates in the xy-plane, perpendicular to the main magnetic field.

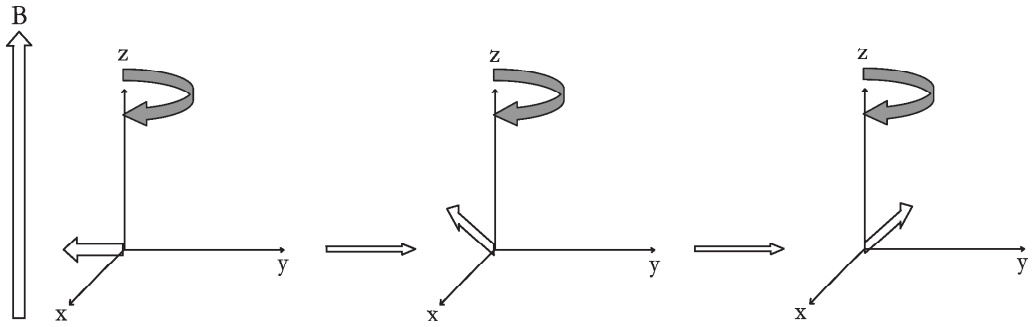

Figure A.2.2.2. Precession of the transverse magnetization. Figure from Spin dynamics: Basics of Nuclear Magnetic Resonance[102].

The precession frequency of the transverse magnetic moment is equal to the precession frequency of the individual spin, i.e. the Larmor frequency.
The macroscopic magnetization components at a time t after the pulse corresponds to[102]:

$$M_y^{nuc} = -M_{eq}^{nuc}\cos\left(\omega^0 t\right)\exp\left\{-t/T_2\right\}$$

$$M_x^{nuc} = M_{eq}^{nuc}\sin\left(\omega^0 t\right)\exp\left\{-t/T_2\right\} \qquad [A.2.2.1]$$

The transverse magnetic moment precesses at the nuclear Larmor frequency ω^0, slowly decaying at the same time

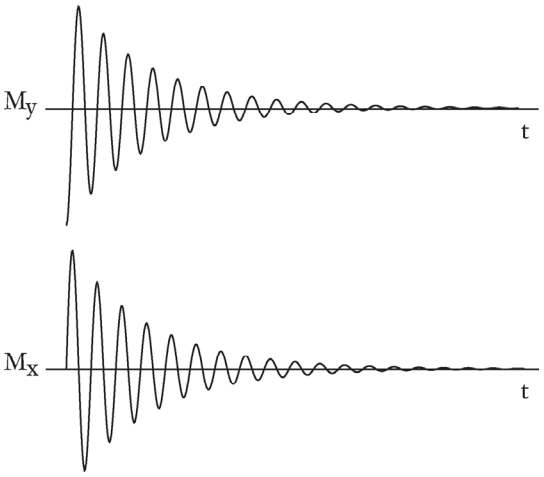

Figure A.2.2.3. Decay and oscillation of the transverse magnetization. Figure from Spin dynamics: Basics of Nuclear Magnetic Resonance[102].

The transverse magnetization decays slowly because it is impossible to maintain exact synchrony between the precessing nuclear magnets. Since the macroscopic magnetic fields fluctuate slightly, the precessing nuclear magnets gradually get out of phase with each other and they will loose coherence.
This decay process is irreversible. Once the transverse magnetization is gone, it cannot be recovered. This is called in *homogenous decay* in NMR jargon.
The time constant T_2 takes into account the homogeneous decay of the precessing macroscopic magnetization. This time constant has various names, the most common being transverse relaxation time constant, coherence dephasing time constant and coherence decay time constant. All that is required is that different spins experience slightly different magnetic fields, so that they precess at slightly different frequencies. This will always be true in a real sample, independent of whether the spins interact with each other.
For NMR of small molecules in liquid, T_2 is typically of the same order of magnitude as T_1, i.e. several seconds. This implies that the nuclear spins execute several tens of million of Larmor precession cycles without losing synchrony. For large molecules or solids, the transverse relaxation time constant T_2 may be as short as milliseconds.

Measuring the T_2-relaxation
Data on transverse relaxation time T_2 for backbone ^{15}N nuclei are essential to study pico-nanosecond as well as micro-millisecond dynamics of proteins. The presence of chemical exchange produces an additional transverse relaxation term R_{ex} to the experimental value $RN(N_{x,y})$, which is proportional on ω_N (Larmor frequency). Thus, the presence of chemical exchange for a residue may be noted by a transverse relaxation rate that is high compared to that of the neighbouring residues. Otherwise, slow dynamics ranging from microsecond to second time scale exchange processes can be studied from ^{15}N relaxation dispersion experiments.; this will be discussed in section A.2.5.

The T_2-relaxation can be measured by a series of HSQC measurements modified so that the magnetization will remain on the nitrogen at different amounts of time in each measurement. The delay time during which the ^{15}N relaxes is specified by a delay, which will be set differently for each determination in the series. The signal intensities are then extracted and by fitting the decline in signal strength over time to a decreasing exponential, the T_2-relaxation can be determined.

A.2.3 Heteronuclear NOE

From the homonuclear case we see that the NOE depends on the rotational correlation time of the internuclear vector of two nuclei, which undergo dipolar relaxation[102]. In principle, the rotational correlation time could be extracted from the intensities of the homonuclear proton-proton NOE experiment. In practice, this is very difficult, because often the exact distances are not known and secondly many protons interact at equal strength since their distances are similar. The heteronuclear NOE between a

proton and directly heteronucleus is much easier to interpret, because the one-bond distance is rather well determined and because this NOE is the dominant relaxation effect on the heteronucleus[89]. ^{15}N NOEs and to a lesser extent[89]. ^{13}C NOEs are used to obtain information on the correlation times of the respective internuclear vectors in biological macromolecules. Usually the steady-state NOE effect is measured. The proton spin is often called I and the heteronuclear spin S. The steady-state NOE enhancement compares the z-magnetization of the S-spin in thermal equilibrium to the z-magnetization of the *S*-spin at equilibrium when the I-spin is saturated[22].

$$NOE\left(\{I\}-S\right)=\frac{\ll S_z \gg_{(I-sat)}}{\ll S_z \gg_{eq}} \qquad \text{[A.2.3.1]}$$

Saturation of the I-spin means that its magnetization vanishes. This can be obtained by applying a large number or random RF-pulses on I. the z-magnetization of the *S*-spin for this case can be calculated as a steady-state solution of the Solomon equation. For the steady-state, the time-derivatives are simply set to 0. Because of the saturation of spin I, we use $< I_z > = 0$

$$0=\begin{pmatrix} -k_I(I) & -k_{NOE} \\ -k_{NOE} & -k_1(S) \end{pmatrix}\begin{pmatrix} -\ll I_z \gg_{eq} \\ \ll S_z \gg_{I-sat} - \ll S_z \gg_{eq} \end{pmatrix} \qquad \text{[A.2.3.2]}$$

solving for $< S_z >_{I\text{-}sat}$ yields

$$0 = k_{NOE} \ll I_z \gg_{eq} -k_I(S)\left(\ll S_z \gg_{I-sat} - \ll S_z \gg_{eq}\right)$$

$$\ll S_z \gg_{I-sat} = \frac{k_{NOE}}{k_1} \ll I_z \gg_{eq} + \ll S_z \gg_{eq} \qquad \text{[A.2.3.3]}$$

and therefore

$$NOE\left(\{I\}-S\right)=\frac{k_{NOE}}{k_1}\frac{\ll I_z \gg_{eq}}{\ll S_z \gg_{eq}}+1=\frac{k_{NOE}}{k_1}\frac{\gamma_I}{\gamma_S}+1 \qquad \text{[A.2.3.4]}$$

k_I is the longitudinal relaxation rate ($1/T_1$). It can also contain contributions from other relaxation mechanisms besides the dipolar 1-bond relaxation. A significant contribution to k_1 for ^{15}N amide nuclei is the relaxation stemming from chemical shift anisotropy (CSA). The Figure shows the behaviour of the amide ^{1}H-^{15}N NOE as a function of the correlation time when this mechanism is taken into account. The effect clearly distinguishes between motions faster or lower than ~ 0.5 s.

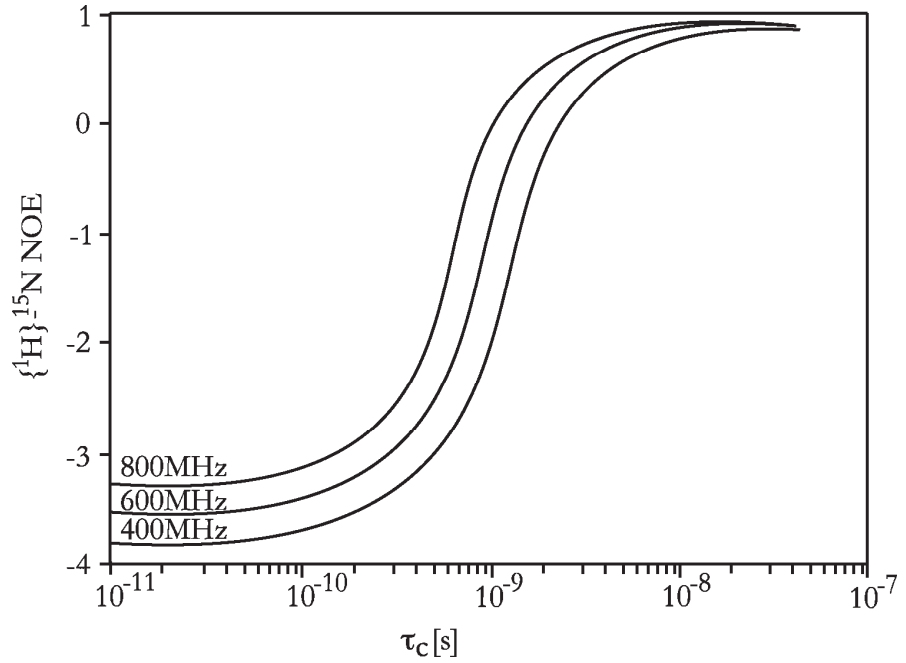

Figure A.2.3.1. Values of the steady-state heteronuclear ^{1}H-^{15}N NOE as a function of correlation time for magnetic field strengths corresponding to 400, 600 and 800 MHz proton Larmor frequency. The effect of 160 ppm chemical shift anisotropy ($\sigma_{\parallel} - \sigma_{\perp}$) on k_1 has been taken into account. Figure from Grzesiek S. 2005[118].

Measuring the heteronuclear NOE

The backbone ^{1}H-^{15}N heteronuclear NOE provides information about the motion of individual N-H bond vectors. Those that undergo motion faster than the overall tumbling of the molecules (i.e. in the pico- nanosecond time scale) show a decreased NOE intensity relative to the average observed for the majority of the residues. Thus, for instance, decreased values are usually found at both N- and C-terminal ends of the protein.

For heteronuclear NOE cross-relaxation rates, $R_S(I_z \rightarrow S_z)$, a single refocused reverse 2D INEPT experiment is used in which proton saturation is achieved during the relaxation T delay prior to the starting 90^{0} ^{15}N pulse. Two different spectra are usually recorded in an interleaved manner with and without proton saturation during 3 - 4 seconds. A long recycle delay is needed in order to ensure complete relaxation of water magnetization at the beginning of each scan.

Values of steady-state NOEs are established from the ratio of peak intensities according to[118]:

$$NOE = \left(I_{saturated} / I_{equilibrium}\right) - 1 \qquad \text{[A.2.3.5]}$$

$R_s(I_z \rightarrow S_z)$ is extracted from combined NOE and $R_s(S_z)$ measurements according to

$$NOE = -9.86 \cdot \left(R_s\left(I_z \rightarrow S_z\right) / R_s\left(S_z\right)\right) \qquad \text{[A.2.3.6]}$$

A.2.4 The correlation time, τ_c.

In a liquid, the random molecular tumbling modulates the magnetic fields exerted by spins on each other. The correlation time of the random fields τ_c corresponds to the rotational time of the molecules. The rotational correlation time is given (roughly) by the average time taken for the molecules to rotate by one radian[12,102]. Generally speaking, small molecules have short rotational correlation times, while large molecules have long rotational times. In reality the τ_c is dependent on the size, shape, and dynamics of the molecule, as well as the bulk physical characteristics of the solvent.

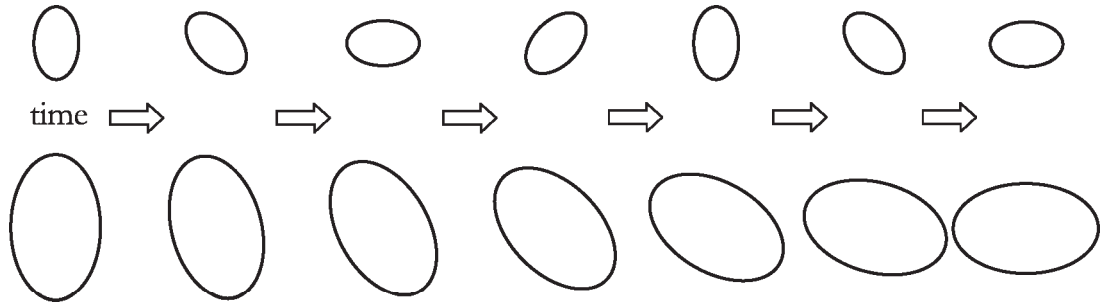

Figure A.2.4.1. Overall tumbling of two macromolecules of different size. In the same time interval, a small molecule rotates more than a large molecule. Figure from Spin dynamics: Basics of Nuclear Magnetic Resonance[102].

The correlation time, τ_c, is dependent on the T_1- and T_2-relaxation rates of the molecule. At very short rotational correlation times, the values of T_1 and T_2 are equal. This is called the extreme narrowing limit. As the correlation time is increased, T_1 passes through a minimum and then increases. The transverse relaxation time constant T_2, on the other hand, continues to decrease. Line width at half-height of the NMR line can be estimated following[102]:

when $T_1 = T_2$

$$linewidth = \frac{\Delta\omega}{2\pi} = \frac{1}{\pi T_1} \qquad \text{[A.2.4.1]}$$

for larger molecules when $T_2 < T_1$

$$linewidth = \frac{\Delta\omega}{2\pi} = \frac{1}{\pi T_2} \qquad \text{[A.2.4.2]}$$

In practice, this means that the NMR peaks get progressively broader as the molecular mass is increased. The experimental relaxation rate, extracted from the line width, is called $1/T_2^*$ ($1/T_2^* = 1/T_2 + 1/T_{2(\text{inhomogeneity})}$).
For many years, this behaviour was believed to set a fundamental limit on the size of molecule that may be studied by solution NMR. However, the development of the TROSY method allows one to overcome the increase in linewidth at large rotational

correlation times.

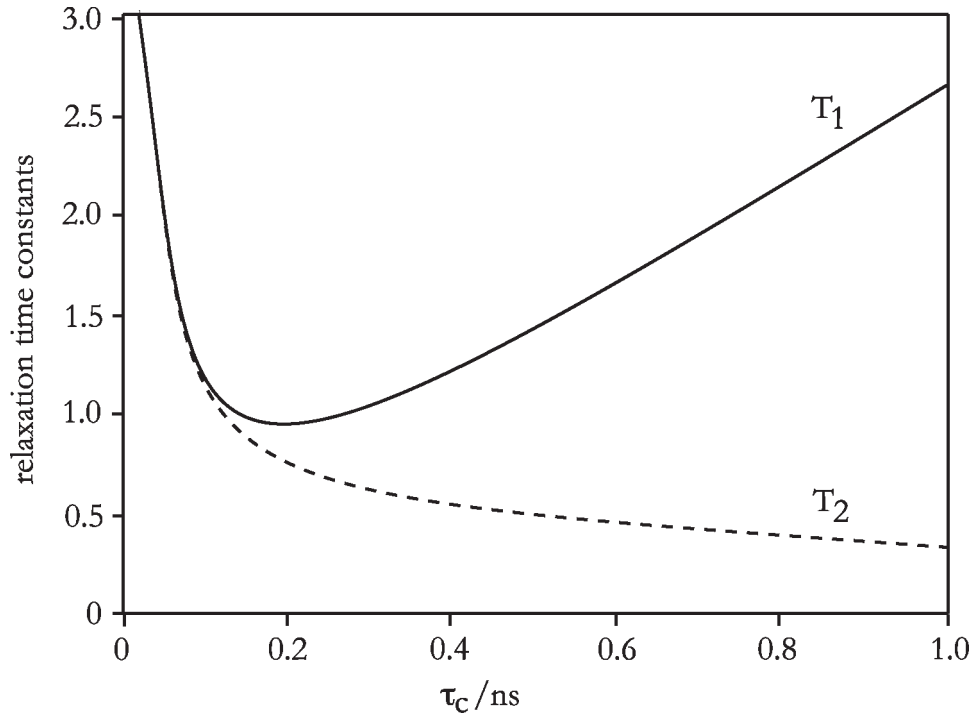

Figure A.2.4.2. Variation of T_1 and T_2 with correlation time, for intramolecular dipole-dipole relaxation. Figure from Grzesiek S. 2005[118].

Oschkinat and co-workers presented a way of estimating the macromolecular τ_c and its molecular weight by analysis of the T_1 and T_2 values[29]. The first step in the analysis is the estimation of the overall correlation times from the trimmed mean T_1/ T_2 ratios. The assumptions underlying such use of the T_1/ T_2 ratio are the following: (i) one assume large order parameter ($S^2 > 0.5$) and rapid internal motion ($\tau_{int} < 100$ ps) for the majority of residues such that the spectral density function is approximated by[118]:

$$J'(\omega) = S^2 \frac{\tau_R}{1 + \omega^2 \tau_R^2} \qquad \text{[A.2.4.3]}$$

which is simply the rigid rotor model scaled by the order parameter; (ii) it is then assumed that for overall correlation times typical of proteins ($\tau_R > 3$ ns), all high-frequency terms ($\omega_H - \omega_N$, ω_H, $\omega_H + \omega_N$) have negligible contributions to the T_1, T_2 and NOE rates. In this manner, the relaxation rates for T_1 and T_2 reduce to:

$$\frac{1}{T_1} = \frac{1}{3}\left[3J(\omega_N)\right]$$

$$\frac{1}{T^2} = \frac{1}{6}\left[3d^2 + c^2\right]\left[4J(0) + 3J(\omega_N)\right] \qquad \text{[A.2.4.4]}$$

and the ratio is given by[103]:

$$\frac{T_1}{T_2} = \frac{\left[4J(0) + 3J(\omega_N)\right]}{6J(\omega_N)} = \frac{7 + 4\left(\omega_N \tau_R\right)^2}{6} \qquad \text{[A.2.4.5]}$$

The overall correlation times are then determined using the trimmed mean T_1/ T_2 ratio, such that after the mean is calculated, all values differing from the mean by one standard deviation are removed and the mean value is recalculated.
In this manner residues undergoing rapid internal motions or conformational exchange and for which the above arguments do not hold are removed from the analysis. The correlation times thus derived are then used to determine which residues have abnormally high R_2 relaxation rates and therefore should be fitted using an additional exchange term, R_{ex}[104]:

$$\frac{1}{T_{2,\exp}} = \frac{1}{T_2} + R_{ex} \qquad \text{[A.2.4.6]}$$

where $T_{2,exp}$ is the experimentally observed relaxation time and T_2 is the component arising from dipolar and CSA mechanisms as discussed earlier. The overall correlation time can then be fitted to experimentally determined molecular weights of different proteins of different shapes and.

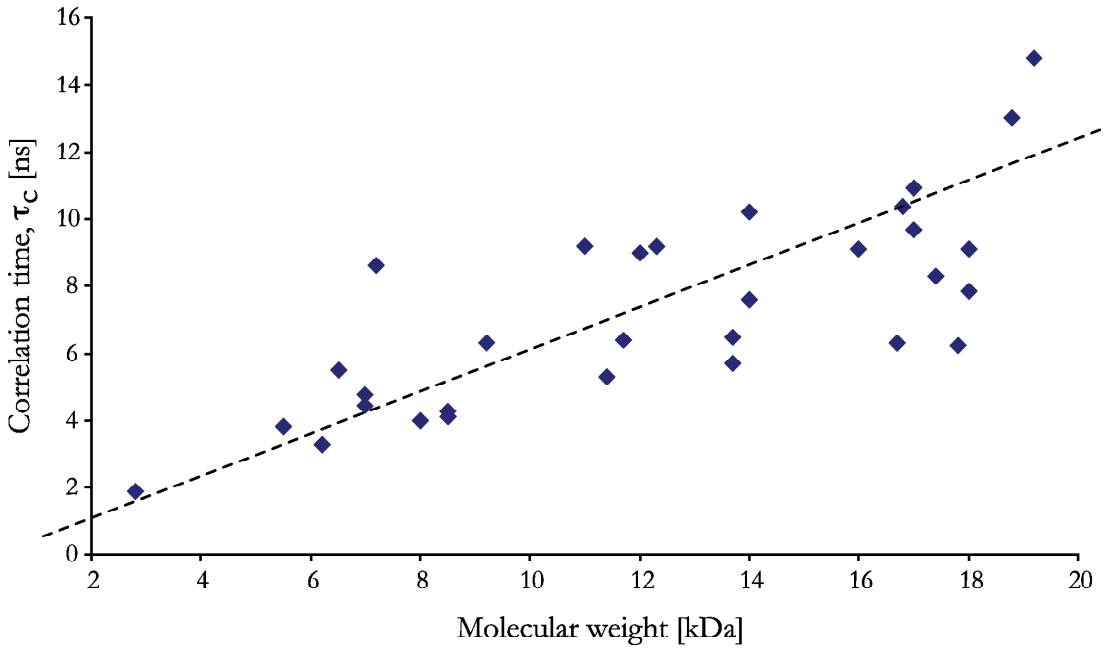

Figure A.2.4.3. Experimentally measured τ_c values for a range of proteins with known molecular weight. Figure from Strauss, H. M.[116]

A.2.5 Carr-Purcell Meiboom-Gill (CPMG)

Folding and unfolding reactions of proteins are of great importance and they often occur on a time-scale, which cannot be probed by NMR in conventional ways. They can however be probed by using so-called Carr-Purcell Meiboom-Gill experiments[105,106], by measuring the contribution to line widths of cross peaks in correlation spectra, T_2-relaxation. They are interpreted in terms of the rates of interconversion between states (kinetics), their chemical shift differences (structure) and their populations (thermodynamics)[107-110]. For exchange rates varying from a few hundred to a few thousand per second, and for populations of the higher-energy (unfolded or intermediate) state greater than about 5 %, this approach provides a sensitive measure of the exchange dynamics[111]. This gives data that are site-specific and in absence of denaturant.
Experimental data has shown that an assumption of a two-site exchange is normally adequate to obtain high quality of the fits of the data and the associated statistical analysis[111,112].

In non-viscous liquids, usually $T_2 = T_1$. But some processes like scalar coupling with quadrupolar nuclei, chemical exchange, interaction with a paramagnetic center, can accelerate the T_2-relaxation such that T_2 becomes shorter than T_1. In principle T_2 can be obtained by measuring the signal width at half-height according to Eq. S1.4.2. However the line width for non-viscous liquids is most often dominated by field inhomogeneity. Fortunately, the dephasing of spins isochromats, resulting from field inhomogeneity, is a reversible process: it can be refocused by using a 180^0 pulse inserted in the center of an evolution time. This is called a Hahn spin-echo[113]. The Hahn echo is constituted first by a 90^0 pulse that flips the magnetization in the XY plane. During the first τ delay, the magnetization evolves according to its chemical shift (and field inhomogeneity). An 180^0 pulse is then applied. This 180^0 pulse inverts the magnetization. (This pulse can be applied along the X or the Y axis). Following the inversion pulse, another τ delay is applied. During this delay, the magnetization refocus: for example, if the nuclei with chemical shift 1 have had the time to precess clockwise by 15^0 degree during the first τ delay, it will also precess by the same amount after the inversion pulse as the identity of the nuclei have not been changed. Similarly, if field inhomogeneity is present for the first τ delay, the same inhomogeneity is present after the inversion pulse and will influence the precession in the same manner allowing to refocus this unwanted effect at the end of the second delay. This will give rise to an echo after the $2*\tau$ delays. The spin-spin relaxation processes will only affect the size of this echo.

At the end of the second delay, the echo will be lined up:

along the Y axis if the 180^0 pulse was applied along the Y axis,
along the -Y axis if the 180^0 pulse was applied along the X axis.

The Carr-Purcell-Meiboom-Gill[105,106] sequence is derived from the Hahn spin-echo

sequence[113]. This sequence is equipped with a "built-in" procedure to self-correct pulse accuracy error.
If the first inversion pulse applied is shorter (e.g. 175^0) than an 180^0 pulse, a systematic error is introduced in the measurement. The echo will form above the XY plane (e.g. 5^0) and therefore the signal will be smaller than expected.
To correct that error, instead of sampling immediately the echo, a third τ delay is introduced, during which, the magnetization evolve as before but slightly above the XY plane.
If the second inversion pulse, also shorter than an 180^0 pulse (e.g. 175^0), is applied, as the magnetization is already above the plane, this shorter inversion pulse will put the magnetization exactly in the XY plane.
At the end of the last *tau* delay, the echo will form exactly in the XY plane, self-correcting the pulse error.

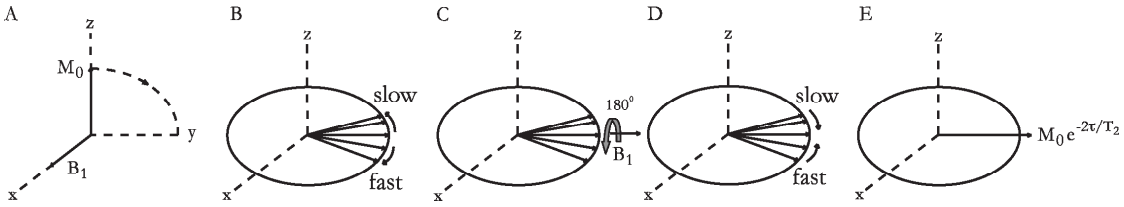

Figure A.2.5.1. The spin-echo effect, (A), a 90^0_x pulse puts M_o into the y' direction, (B), the spin isochromats fan out, (C), a 180^0_y pulse interchanges slow and fast spins at time τ, (D), refocusing occurs, (E), the echo at time 2τ. Figure from Grzesiek S. 2005[118].

The process of refocusing by the 180^0 pulse can be repeated many times. The amplitudes of successive echoes decay exponentially and the true value of T_2 can be found from the envelope of the echoes.

A.2.6 Relaxation changes and molecular interactions

Monitoring the transverse relaxation (T_2) of the ligand signals is one of the most common methods to detect ligand binding by NMR. The transverse relaxation rate ($1/T_2$) of a small molecule tumbling rapidly in solution is typically longer than those of the protein molecules and ligands bound to protein, which are tumbling much slower. The linewidths in the spectra are directly dependent on the tumbling rate, τ_c, and by linewidth analysis the interaction between the protein-ligand, protein-protein interaction can be determined. The interaction can be done more elaborate by using a T_2 relaxation filter[11-13,32]. By this method the signals of the bound ligand can be selectively filtered out. In this method, two relaxation-edited spectra are obtained, one with and one without the ligand, upon differencing these spectra the binding ligand can be obtained.

A.3 Residual dipolar couplings (RDCs)

While NMR spectroscopy is now successfully established as the most important technique for the high resolution structure determination of small to medium sized, compact macromolecules in the solution state[20,32,54], the method has some notable limitations for more complex molecular systems.
The basic experimental parameter used for the determination of molecular structure (Nuclear Overhauser Effect, NOE) becomes difficult to measure in large protonated molecules due to prohibitive relaxation effects, making the determination of structure beyond 30 kDa unrealistic using classical techniques. Moreover, modular or elongated proteins, and large RNA superstructures, encounter the serious problem of ill-defined relative orientation of different domains, due to inadequate local structural information at interfacial or hinge regions. The relative orientation of different domains is however known to be closely correlated to physiological function while the characterization of the exact nature of molecular interaction in reaction complexes clearly holds the key to understanding macromolecular function.
The last five years has seen a rapid acceleration in the search for viable, alternative sources of structural information for the resolution of long-range orientation in systems of more complex geometry[55].
In particular weak alignment of proteins prevents complete averaging of the dipolar interaction, while retaining the solution properties necessary for high resolution NMR. This alignment can exist naturally, due to the paramagnetic properties of the molecule[56], or can, more generally, be induced by solvation in liquid crystal media[57]. The residual dipolar coupling (RDC) measured under these conditions provides geometric information relative to the common alignment frame.
The degeneracy of the RDCs can be raised if we can measure more couplings in a domain of known structure, and whose relative orientation in the domain is known. There are now four equivalent orientations of the differently valued couplings, which are in agreement with measure values. This four-fold degeneracy is inherent to the orientation of any three-dimensional structure relative to a molecular alignment tensor, and derives from simple symmetry operations (180° rotations around A_{xx}, A_{yy} and A_{zz}). Despite this inherent four-fold orientational degeneracy, the ability to determine domain orientation is a very powerful complement to classical structure determination and forms the basis of many recent studies of the molecular architecture of multidomain systems[58-60], and protein-ligand complexes[61,62].
The use of residual dipolar couplings have been around for a long time, but prior to 1995 the measurement of residual dipolar couplings (RDCs) was confined to small molecules[12,63]. Only in recent years the technique has entered macromolecular solution nuclear magnetic resonance (NMR) spectroscopy. Since then the number of pulse sequences and alignment media required has been growing steadily, and new applications for residual dipolar couplings are continually being reported. Residual dipolar couplings are used in structure determinations, protein fold recognition of folded and denatured proteins and for insight into dynamic mechanisms[55,64,65].

A.3.1 The dipolar interaction

A dipolar interaction arises from the interaction of a pair of nuclear magnetic moments. A nucleus will experience a local magnetic field from a neighbouring nucleus depending if the nucleus' magnetic dipole reinforces or opposes the static magnetic field B_0. The strength of the local field induced in a neighbouring atom is dependent on several factors, the nature of the nuclei (via its magnetogyric ratios), the distance between the nuclei (r_{xy}) and the angle θ, which the internuclear vector makes with the static external magnetic field.

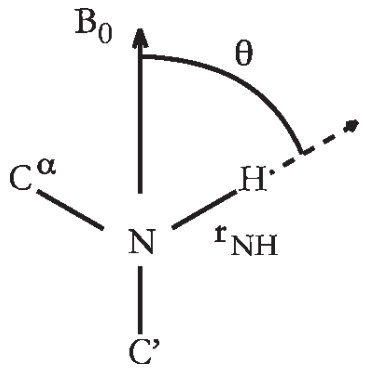

Figure A.3.1.1. Diagram showing two parameters which define the size of the dipolar coupling between the N and H^N atoms of a protein backbone: r_{NH}, the N-H^N bond length and the angle, θ, between the N-H^N bond and the static magnetic field, B_0.

More specifically the dipolar coupling between two nuclei A and B is given by Eq A.1.6.1.1. Where S is the generalised order parameter for internal motion of the AB vector, μ_0 is the magnetic permeability of vacuum, γ_A and γ_B are the magnetogyric ratios of A and B, h is Planck's constant, r_{AB} is the distance between A and B, A_a and A_r are the unit less axial and rhombic components of the alignment tensor A, and θ and ϕ are the cylindrical coordinates describing the orientation of the vector AB in the principal axis system of the alignment tensor[66].

$$D_{AB} = \frac{-S\mu_0\gamma_A\gamma_B h\left[A_a\left(3\cos^2\theta - 1\right) + \frac{3}{2}A_r\left(\sin^2\theta\cos 2\phi\right)\right]}{16\pi^3 r_{AB}^3} \qquad [A.3.1.1]$$

This equation simplifies to[54]

$$D_{AB} = D_a\left[\left(3\cos^2\theta - 1\right) + \frac{3}{2}R\left(\sin^2\theta\cos 2\phi\right)\right] \qquad [A.3.1.2]$$

where

$$D_a = \frac{-S\mu_0\gamma_A\gamma_B hA_a}{16\pi^3 r_{AB}^3} = \frac{1}{3}\left[D_{zz} - \frac{\left(D_{xx} + D_{yy}\right)}{2} \right] \qquad [A.3.1.3]$$

$$D_r = \frac{-S\mu_0\gamma_A\gamma_B hA_r}{16\pi^3 r_{AB}^3} = \frac{\left(D_{xx} - D_{yy}\right)}{2} \qquad [A.3.1.4]$$

$$R = \frac{A_r}{A_a} \qquad [A.3.1.5]$$

In these equations D_a and D_r are the magnitudes of the axial and rhombic components of the alignment tensor, respectively, and R is known as the rhombicity.
The alignment tensor, which is used to describe the magnitude and orientation of the alignment of the molecule, can be expressed as an order matrix, S. This is both traceless ($S_{xx} + S_{yy} + S_{zz} = 0$) and symmetric ($S_{ij} = S_{ji}$) [64,67]. It therefore contains only five independent elements, which correspond to the magnitude of the axial, D_a, and rhombic, D_r, components of the alignment tensor, as well as three Euler angles (α, β, γ) describing the orientation of the principle component of the tensor with respect to the molecular frame.
In solid-state NMR large dipolar couplings, order of kHz, can be measured and these provide important structural information. However, in the case of solution-state NMR the solution is isotropic and the dipolar couplings are averaged to zero. Producing small net alignment of molecules creates residual dipolar couplings, which are observable.
Most diamagnetic molecules will have a tendency to align with a magnetic field due to their magnetic susceptibility anisotropy. In proteins the magnetic susceptibility is made anisotropic through ring currents from aromatic residues such as Histidine, Phenylalanine, Tryptophan and Tyrosine. However, the effects from these residues often cancel one another out and the magnetic susceptibility anisotropy is generally very small in proteins.

A.3.2 Inducing alignment

A breakthrough was the alignment of a protein in liquid crystalline medium[57]. Since the first report in 1997, it has become possible to dissolve proteins in a variety of non-isotropic environments, such as liquid crystalline solutions or anisotropically strained gel matrices. These media all result in a weak net alignment of the protein with respect to the static magnetic field, by biasing the tumbling of the protein due to steric or electrostatic effects. This makes the measurement of small residual dipolar couplings of the order of 10 – 20 Hz possible, while still leaving the molecules tumbling rapidly enough to produce sharp solution-state NMR spectra.
All work involving RDCs assumes that the alignment of the biomolecule occurs

without altering its structure and that there is no direct contact or binding between the molecule and the alignment medium. This is supported by the fact that chemical shifts are the same in the isotropic and anisotropic medium. Some cases in which interactions between the protein and the alignment medium do occur have been reported, but are rare[68,69]. If the protein being aligned is charged, then the alignment medium must either be of the same charge or high concentrations of salt needs to be present to counter the electrostatic effects[70,71]. Nonetheless the possibility of the alignment medium affecting the structure of the protein continues to be a concern and several studies have been conducted which attempt to show that alignment media only disturb the protein structure on a non-significant way.
In many cases the alignment is achieved purely by static interactions, such as with neutral bicelles or *n*-alkyl(PEG)/*n*-alkyl alcohol mixtures. This has been verified by producing models, which mimic the steric interaction between a protein and the aligning medium and will predict RDCs on this basis. These programs show good agreement with experimental data from neutral alignment media, indicating that no other significant interaction occur[72-75].

A.3.3 Determination of the alignment tensor

In general, the alignment tensor of a given protein is not known *a priori*. In addition, any single measured RDC value does not uniquely define a vector orientation. Rather it defines two cones, of opposite sign, about the z-axis[76]. If the structure of the molecule is known then the cones can be narrowed down to exact positions due to the constraints of the molecular structure. The molecule can be rotated to obtain the best fit alignment tensor[77]. Alternatively an order matrix analysis can be conducted using singular value decomposition, provided a minimum of five RDCs have been measured[67].

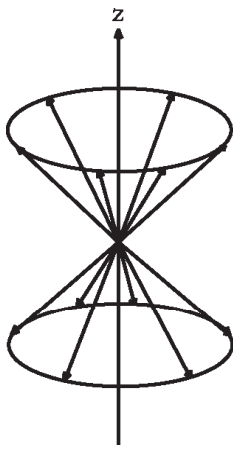

Figure A.3.3.1. Diagram showing the degeneracy of the directions defined by a single RDC value.

In situations where the structure is not known, the magnitude of the axial component of the alignment tensor and the rhombicity can be estimated from the distribution of

the normalised RDCs[78]. If more than one type of RDC is available, RDC values are normalised, usually to the N – H^N RDCs, using a factor of $(\gamma_N\gamma_H r^3_{PQ})/(\gamma_P\gamma_Q r^3_{NH})$. A histogram of these values is plotted and provided the distribution of the vector is relatively uniform, this should resemble a chemical shift anisotropy powder pattern. Values for D_a and R can then be extracted by noting that the minimum RDC occurs at $\theta = 90^0$, $\phi = 90^0$, the maximum RDC occurs at $\theta = 0^0$ and the highest probability RDC occurs at $\theta = 90^0$, $\phi = 0^0$. From equations A.1.6.1.2-5 it follows that

$$D_a = \frac{1}{2} D_{\max} \qquad \text{[A.3.3.1]}$$

$$R = \left[\left(\frac{-D_{\min}}{D_a} \right) - 1 \right] \left(\frac{3}{2} \right) \qquad \text{[A.3.3.2}$$

An extension to this method has been made which uses a fitting procedure to optimise the shape of the histogram and extract values for D_a and R[79]. The authors show that although this method is more accurate than that proposed by Clore *et al.*[54], it is nevertheless equally limited by the restricted sampling of orientational space[60]. More recently, a maximum likelihood technique has been proposed for the estimation of D_a and R which is particularly advantageous when using small datasets or if the distribution of vectors is particularly non-uniform[80].

A theoretical approach has been described which makes it possible to analyse how well a set of RDCs samples the tensorial space and therefore how likely it is that the alignment tensor can be accurately defined[81]. This analysis shows that N-H^N vectors tend to sample the available orientations very badly[81]. This is easily rationalised by observing that the N-H^N moiety is important in fixing secondary structure via hydrogen bonds and therefore has a strong tendency to point in the same regular directions in any single element of secondary structure (e.g. in a helix all N-H^N bonds point nearly parallel to the axis of the helix). By contrast the C^α – C' and N-C' bond vectors tend to sample orientational space better[81], though the C^α-C' and N-C' RDCs often carry larger experimental errors due to their smaller size.

Data from two different alignment media can reduce the degeneracy problem. The cones for a vector from two different orientations overlap, and so the orientation of the vector is reduced to, maximally, eight points of intersection of the cones[76]. By analysing simple correlation plots of RDCs from two different alignment media, it is possible to determine the relative orientation of the two alignment tensors[82]. With as many as five or more data sets, it is possible to reconstruct all vector orientations completely without any previous knowledge of the structure[83], however the problem of a two-fold degeneracy always remains[64].

A.3.4 Alignment media

Ever since Tjandra and Bax published their report of measuring RDCs in a protein suspended in a dilute liquid crystalline medium[66], a growing number of media for the alignment of biological macromolecules have become available.
A widely used alignment medium, which was in used in this work, is Pf1 filamentous bacteriophage. It consists of rod shaped particles of ~60 Å diameter and ~20,000 Å length and has a negatively charged surface with a pI of ~4.0. It aligns in a magnetic field and induces alignment in proteins, DNA and RNA[84]. By varying the Pf1 concentration of the solution the degree of alignment can be altered[84]. Addition of salts is possible[85], and can enable the alignment of positively charged proteins which would otherwise stick to the surface of the phage. Alternatively, working at very low (0 – 4 mg/ml) concentrations of phage has also made the measurement of RDCs in a positively charged protein possible[69]. In recent years Pf1 phages appears to have become one of the main aligning media used[86-91]. As an alternative to Pf1 phage, Fd phage and tobacco mosaic virus have been used to align proteins[20].

A.3.5 Domain orientations and protein complex structures

Defining the orientation of two protein domains relative to one another can be difficult using conventional NMR techniques because there is often a lack of sufficient numbers of inter-domain NOE restraints to define the orientation fully. However, using RDCs in a similar way as for normal structure determinations can help yield the correct orientation, as has been found to be the case in a number of systems[68,92].
In some proteins the crystal packing is thought to distort the inter-domain orientation in crystal structures away from that in the solution structure. By measuring RDCs for both domains and fitting these separate alignment tensors, the correct solution orientation can be found[60,93,94]. Alternatively the two domains can be subjected to a simulated annealing protocol to determine the correct orientation[95,96].
RDCs can also be an indicator for inter-domain dynamics. For example, if the magnitude of the alignment differs markedly between the two domains, this may be because one is more strongly aligned than the other and there is substantial inter-domain motion present[68,79,97,98]. Similar processes are also possible with protein-protein or protein-ligand complexes[93,99-101].

A.3.6 RDC NMR experiments

A large number of two and three-dimensional pulse sequences for the measurement of RDCs have been reported in the literature. The most commonly measured RDC is the $N\text{-}H^N$ RDC. This is generally measured by recording an HSQC-type spectrum in which the nitrogen decoupling has been removed in the indirect dimension. To avoid overcrowding of spectral peaks two approaches are possible; (i) the two components of each doublet can be separated into two different spectra. This can either be done by

recording an in-phase and an anti-phase spectrum, which are added and subtracted to obtain the individual components, IPAP method[114], or by a spin-state selective approach, S^3E method[115]. (ii) The dipolar coupling can also be measured by recording a regular HSQC-spectrum and an HSQC trosy-type spectrum. This will effectively give half the coupling that can just be multiplied by two to give the true value. This approach is especially well suited for larger macromolecular systems (P. Schmieder, personal communication).

Pulse sequences for other RDCs such as C^{α}-H^{α}, C^{α}-C', C'-N and C'-H^N have been comprehensively reviewed by Bax *et al*.[64]. These tend to fall into two categories of either being decoupled versions of other experiments, or (particularly for C-H RDCs) being J-modulated experiments[64].

A.4 RDCs of MalF-P2 in presence and absence of MalE

RDCs for MalF-P2 in absence and presence of MalE (molar ratio 1:1). All residues are listed according to the full length MalF primary sequence, specifying residue type, nucleus, the observed residual dipolar coupling (D_OBS), the back calculated reidual coupling (D) using the PDB file 2r6g, and the difference in coupling (D_DIFF), respectively.

H^NN all (in absence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
100	LEU	H	100	LEU	N	1,6442	1,399	0,2452
101	THR	H	101	THR	N	4,9807	7,409	-2,4283
102	PHE	H	102	PHE	N	3,0967	5,807	-2,7106
105	ALA	H	105	ALA	N	4,6432	7,090	-2,4475
106	GLN	H	106	GLN	N	-5,8213	3,887	-9,7085
107	GLU	H	107	GLU	N	4,3539	0,602	3,7517
113	SER	H	113	SER	N	-2,4860	-12,45	9,9696
115	GLN	H	115	GLN	N	-8,6411	-4,078	-4,563
116	ALA	H	116	ALA	N	-2,3500	4,338	-6,6888
117	GLY	H	117	GLY	N	-9,9317	-1,471	-8,4599
118	LYS	H	118	LYS	N	-16,798	0,221	-17,020
119	THR	H	119	THR	N	12,250	1,414	10,836
121	ASN	H	121	ASN	N	12,327	7,980	4,347
123	GLY	H	123	GLY	N	-22,146	-12,71	-9,4292
125	TYR	H	125	TYR	N	-16,355	-4,475	-11,879
127	ALA	H	127	ALA	N	13,213	6,808	6,4059
130	GLU	H	130	GLU	N	14,735	6,004	8,7313
132	GLN	H	132	GLN	N	-19,644	-12,56	-7,0762
134	ALA	H	134	ALA	N	-11,337	-8,822	-2,5147
135	LEU	H	135	LEU	N	-12,869	-5,879	-6,9897
137	ASP	H	137	ASP	N	-6,3875	-4,278	-2,1092
140	THR	H	140	THR	N	11,275	3,886	7,3892
141	GLY	H	141	GLY	N	-1,3079	3,054	-4,3624
144	TYR	H	144	TYR	N	-18,523	-11,03	-7,4837
145	LEU	H	145	LEU	N	-13,928	-9,756	-4,1717
147	ASP	H	147	ASP	N	5,9944	3,918	2,0763
148	ALA	H	148	ALA	N	-9,1381	-1,102	-8,0355
150	LYS	H	150	LYS	N	-12,404	-8,700	-3,7041
151	PHE	H	151	PHE	N	-0,6638	3,966	-4,6302
152	GLY	H	152	GLY	N	9,4855	0,826	8,6587
153	GLY	H	153	GLY	N	-23,700	-11,72	-11,975
156	LYS	H	156	LYS	N	-0,7281	-7,565	6,8372
157	LEU	H	157	LEU	N	-2,2549	-3,220	0,9655
158	GLN	H	158	GLN	N	7,9006	0,423	7,4771
159	LEU	H	159	LEU	N	14,634	8,912	5,7217
161	GLU	H	161	GLU	N	-14,116	-10,10	-4,0155
164	ALA	H	164	ALA	N	0,2905	-3,056	3,3469
168	GLY	H	168	GLY	N	5,4430	6,426	-0,9835
170	ALA	H	170	ALA	N	14,242	3,270	10,972

177	THR	H	177	THR	N	-2,3760	2,900	-5,2763
178	GLN	H	178	GLN	N	9,9985	6,290	3,7083
179	ASN	H	179	ASN	N	9,7574	1,514	8,2433
180	ARG	H	180	ARG	N	-16,617	-9,923	-6,6936
182	ALA	H	182	ALA	N	-11,645	-0,783	-10,861
183	LEU	H	183	LEU	N	-21,902	-9,697	-12,204
184	SER	H	184	SER	N	-11,967	-1,608	-10,359
185	ASP	H	185	ASP	N	-0,1038	3,647	-3,7512
186	ILE	H	186	ILE	N	-14,012	-9,650	-4,3618
188	ALA	H	188	ALA	N	-7,0637	-11,52	4,4635
190	LEU	H	190	LEU	N	7,0180	8,716	-1,6985
192	ASP	H	192	ASP	N	-14,247	-12,43	-1,8116
193	GLY	H	193	GLY	N	-1,7196	-9,435	7,7156
196	VAL	H	196	VAL	N	1,7319	2,484	-0,7527
198	MET	H	198	MET	N	3,9967	0,292	3,7046
199	SER	H	199	SER	N	3,5282	7,294	-3,7659
200	SER	H	200	SER	N	-2,0286	-2,435	0,4073
201	LEU	H	201	LEU	N	-11,421	-6,313	-5,1078
204	PHE	H	204	PHE	N	6,1440	5,575	0,5686
205	SER	H	205	SER	N	11,625	3,188	8,4367
206	GLY	H	206	GLY	N	3,5207	1,984	1,5358
210	LEU	H	210	LEU	N	7,6423	-8,991	16,633
211	TYR	H	211	TYR	N	8,2085	1,934	6,2744
213	LEU	H	213	LEU	N	12,208	5,258	6,9507
214	ASP	H	214	ASP	N	2,2400	3,661	-1,4215
217	GLY	H	217	GLY	N	8,3815	3,368	5,0134
219	LEU	H	219	LEU	N	3,3637	-1,431	4,7955
220	THR	H	220	THR	N	-5,9808	7,433	-13,413
221	ASN	H	221	ASN	N	8,6906	8,875	-0,1844
222	ASN	H	222	ASN	N	10,208	-9,517	19,725
227	LYS	H	227	LYS	N	11,614	3,557	8,0568
228	TYR	H	228	TYR	N	2,4205	3,183	-0,7631
232	ASN	H	232	ASN	N	4,3181	-10,86	15,185
237	TYR	H	237	TYR	N	-6,2070	-0,589	-5,6174
238	GLN	H	238	GLN	N	-16,196	3,457	-19,654
239	SER	H	239	SER	N	-9,0268	-5,465	-3,5610
240	ILE	H	240	ILE	N	13,990	8,851	5,1388
241	THR	H	241	THR	N	7,6287	6,370	1,2582
245	ASN	H	245	ASN	N	-1,5354	1,170	-2,7054
247	GLY	H	247	GLY	N	-5,5358	1,700	-7,2362
248	ASP	H	248	ASP	N	8,8414	-4,054	12,895
250	LYS	H	250	LYS	N	-13,451	-2,964	-10,486
254	GLY	H	254	GLY	N	-5,2428	-4,147	-1,0949
255	TYR	H	255	TYR	N	4,2847	2,419	1,8657
257	VAL	H	257	VAL	N	-8,3383	-3,406	-4,9314
260	GLY	H	260	GLY	N	10,730	6,945	3,7853
268	PHE	H	268	PHE	N	-1,4105	-6,721	5,3113
269	THR	H	269	THR	N	-2,8655	6,627	-9,4932
271	GLU	H	271	GLU	N	1,8345	-10,40	12,243
272	GLY	H	272	GLY	N	-4,2452	1,575	-5,8210
273	ILE	H	273	ILE	N	-8,9972	-11,46	2,4667
274	GLN	H	274	GLN	N	2,2227	-6,362	8,5855
275	LYS	H	275	LYS	N	3,6060	-1,031	4,6379

H^NN domain 1 (in absence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
100	LEU	H	100	LEU	N	1,6442	-1,449	3,0939
101	THR	H	101	THR	N	4,9807	6,688	-1,7075
102	PHE	H	102	PHE	N	3,0967	-0,3	3,3971
105	ALA	H	105	ALA	N	4,6432	0,32	4,3226
106	GLN	H	106	GLN	N	-5,8213	-1,012	-4,8085
107	GLU	H	107	GLU	N	4,3539	5,386	-1,0320
113	SER	H	113	SER	N	-2,4860	-0,427	-2,0583
210	LEU	H	210	LEU	N	7,6423	6,801	0,8404
211	TYR	H	211	TYR	N	8,2085	6,779	1,4287
213	LEU	H	213	LEU	N	12,208	6,665	5,5438
214	ASP	H	214	ASP	N	2,2400	-2,7	4,9403
217	GLY	H	217	GLY	N	8,3815	6,265	2,1163
219	LEU	H	219	LEU	N	3,3637	6,163	-2,7992
220	THR	H	220	THR	N	-5,9808	5,774	11,755
221	ASN	H	221	ASN	N	8,6906	6,416	2,2741
222	ASN	H	222	ASN	N	10,208	5,564	4,6445
227	LYS	H	227	LYS	N	11,614	6,546	5,0674
228	TYR	H	228	TYR	N	2,4200	-1,152	3,5728
232	ASN	H	232	ASN	N	4,3180	-5,392	9,7104
237	TYR	H	237	TYR	N	-6,2070	-1,556	-4,6509
238	GLN	H	238	GLN	N	-16,190	-9,749	-6,4472
239	SER	H	239	SER	N	-9,0268	-8,655	-0,3710
240	ILE	H	240	ILE	N	13,990	6,531	7,4588
241	THR	H	241	THR	N	7,6287	5,977	1,6515
245	ASN	H	245	ASN	N	-1,5354	-12,502	10,966
247	GLY	H	247	GLY	N	-5,5358	-2,289	-3,2467
248	ASP	H	248	ASP	N	8,8414	6,689	2,1523
250	LYS	H	250	LYS	N	-13,451	-11,51	-1,9324
254	GLY	H	254	GLY	N	-5,2428	0,38	-5,6234
255	TYR	H	255	TYR	N	4,2847	6,527	-2,2432
257	VAL	H	257	VAL	N	-8,3383	-7,067	-1,2709

H^NN domain 2 (in absence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
117	GLY	H	117	GLY	N	-9,9317	-4,654	-5,2772
118	LYS	H	118	LYS	N	-16,798	4,891	-21,689
119	THR	H	119	THR	N	12,250	6,090	6,1608
121	ASN	H	121	ASN	N	12,327	9,374	2,9527
123	GLY	H	123	GLY	N	-22,146	-18,49	-3,6495
125	TYR	H	125	TYR	N	-16,355	-5,697	-10,657
127	ALA	H	127	ALA	N	13,213	10,98	2,2324
130	GLU	H	130	GLU	N	14,735	7,362	7,3730
132	GLN	H	132	GLN	N	-19,644	-16,81	-2,8275
134	ALA	H	134	ALA	N	-11,337	-14,61	3,2790
135	LEU	H	135	LEU	N	-12,869	-9,799	-3,0692
137	ASP	H	137	ASP	N	-6,3875	-6,259	-0,1281
140	THR	H	140	THR	N	11,275	7,335	3,9398
141	GLY	H	141	GLY	N	-1,3079	2,649	-3,9574

144	TYR	H	144	TYR	N	-18,523	-17,134	-1,3887
145	LEU	H	145	LEU	N	-13,928	-16,84	2,9172
147	ASP	H	147	ASP	N	5,9944	3,219	2,7749
148	ALA	H	148	ALA	N	-9,1381	-2,766	-6,3720
150	LYS	H	150	LYS	N	-12,404	-10,90	-1,5030
151	PHE	H	151	PHE	N	-0,6638	0,908	-1,5724
152	GLY	H	152	GLY	N	9,4855	5,440	4,0452
153	GLY	H	153	GLY	N	-23,700	-17,16	-6,5335
156	LYS	H	156	LYS	N	-0,7281	-4,697	3,9696
157	LEU	H	157	LEU	N	-2,2549	-1,329	-0,9258
158	GLN	H	158	GLN	N	7,9006	4,539	3,3609
159	LEU	H	159	LEU	N	14,634	8,678	5,9557
161	GLU	H	161	GLU	N	-14,116	-16,20	2,0869
164	ALA	H	164	ALA	N	0,2905	1,294	-1,0043
168	GLY	H	168	GLY	N	5,4430	10,73	-5,2938
170	ALA	H	170	ALA	N	14,242	8,996	5,2461
177	THR	H	177	THR	N	-2,3760	3,423	-5,7999
178	GLN	H	178	GLN	N	9,9985	10,07	-0,0748
179	ASN	H	179	ASN	N	9,7574	7,803	1,9536
180	ARG	H	180	ARG	N	-16,617	-17,47	0,8599
182	ALA	H	182	ALA	N	-11,645	-5,659	-5,9857
183	LEU	H	183	LEU	N	-21,902	-17,35	-4,5496
184	SER	H	184	SER	N	-11,967	-5,770	-6,1977
185	ASP	H	185	ASP	N	-0,1038	0,162	-0,2668
186	ILE	H	186	ILE	N	-14,012	-14,25	0,2418
188	ALA	H	188	ALA	N	-7,0637	-14,79	7,7304
190	LEU	H	190	LEU	N	7,0180	10,10	-3,0852
192	ASP	H	192	ASP	N	-14,247	-17,37	3,1284
193	GLY	H	193	GLY	N	-1,7196	-14,55	12,830
196	VAL	H	196	VAL	N	1,7319	5,886	-4,1545
198	MET	H	198	MET	N	3,9967	6,418	-2,4213
199	SER	H	199	SER	N	3,5282	11,02	-7,4939
200	SER	H	200	SER	N	-2,0286	-5,033	3,0049
201	LEU	H	201	LEU	N	-11,421	-7,742	-3,6790
204	PHE	H	204	PHE	N	6,1440	8,712	-2,5683
205	SER	H	205	SER	N	11,625	7,046	4,5793
206	GLY	H	206	GLY	N	3,5207	7,810	-4,2896

HαCα all (in absence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
99	GLN	HA	99	GLN	CA	-0,5500	-2,627	2,0778
100	LEU	HA	100	LEU	CA	-1,0142	-0,164	-0,8498
101	THR	HA	101	THR	CA	-33,812	-16,67	-17,135
102	PHE	HA	102	PHE	CA	31,610	1,872	29,737
103	GLU	HA	103	GLU	CA	-24,517	18,78	-43,302
104	ARG	HA	104	ARG	CA	-8,5401	-11,60	3,0602
105	ALA	HA	105	ALA	CA	20,058	2,722	17,335
106	GLN	HA	106	GLN	CA	14,003	-0,361	14,364
107	GLU	HA	107	GLU	CA	-12,843	20,77	-33,615
108	VAL	HA	108	VAL	CA	-28,359	-18,08	-10,272
112	ARG	HA	112	ARG	CA	-28,895	14,97	-43,867
113	SER	HA	113	SER	CA	3,5228	22,13	-18,614

114	TRP	HA	114	TRP	CA	6,4419	21,36	-14,920
115	GLN	HA	115	GLN	CA	14,578	7,468	7,1094
118	LYS	HA	118	LYS	CA	-3,6559	12,68	-16,338
119	THR	HA	119	THR	CA	-42,733	-2,473	-40,259
120	TYR	HA	120	TYR	CA	-48,703	-4,201	-44,502
121	ASN	HA	121	ASN	CA	10,949	2,091	8,8587
127	ALA	HA	127	ALA	CA	-50,200	-18,37	-31,824
130	GLU	HA	130	GLU	CA	-22,655	-17,21	-5,4450
132	GLN	HA	132	GLN	CA	15,085	9,521	5,5641
133	SER	HA	133	SER	CA	52,614	20,29	32,321
134	ALA	HA	134	ALA	CA	39,265	19,46	19,796
135	LEU	HA	135	LEU	CA	38,252	16,77	21,477
136	SER	HA	136	SER	CA	29,204	10,00	19,200
137	ASP	HA	137	ASP	CA	-3,7417	-20,07	16,335
139	GLU	HA	139	GLU	CA	23,911	5,171	18,740
140	THR	HA	140	THR	CA	-29,459	-2,439	-27,020
142	LYS	HA	142	LYS	CA	-11,963	-16,10	4,1401
143	ASN	HA	143	ASN	CA	14,327	-1,245	15,572
144	TYR	HA	144	TYR	CA	27,688	7,956	19,732
146	SER	HA	146	SER	CA	32,880	8,222	24,657
147	ASP	HA	147	ASP	CA	-23,154	-22,84	-0,3128
148	ALA	HA	148	ALA	CA	41,522	24,37	17,147
149	PHE	HA	149	PHE	CA	22,204	22,78	-0,5832
151	PHE	HA	151	PHE	CA	-21,752	-2,413	-19,339
154	GLU	HA	154	GLU	CA	10,189	16,80	-6,6125
155	GLN	HA	155	GLN	CA	32,250	16,51	15,732
157	LEU	HA	157	LEU	CA	2,4039	22,13	-19,735
158	GLN	HA	158	GLN	CA	-21,686	-10,36	-11,319
159	LEU	HA	159	LEU	CA	-7,6420	-8,226	0,5847
161	GLU	HA	161	GLU	CA	46,384	15,66	30,722
162	THR	HA	162	THR	CA	45,347	15,57	29,775
163	THR	HA	163	THR	CA	-0,9199	3,400	-4,3207
164	ALA	HA	164	ALA	CA	-14,192	2,751	-16,944
166	PRO	HA	166	PRO	CA	31,852	23,73	8,1183
167	GLU	HA	167	GLU	CA	-7,6240	-2,634	-4,9893
169	GLU	HA	169	GLU	CA	33,201	-3,826	37,028
170	ALA	HA	170	ALA	CA	3,0521	12,28	-9,2295
172	ASN	HA	172	ASN	CA	-2,2039	1,834	-4,0379
173	LEU	HA	173	LEU	CA	-0,1915	5,414	-5,6057
176	ILE	HA	176	ILE	CA	52,755	24,59	28,165
177	THR	HA	177	THR	CA	-6,4409	-11,54	5,1062
178	GLN	HA	178	GLN	CA	-18,553	-7,397	-11,155
179	ASN	HA	179	ASN	CA	-41,422	0,086	-41,509
180	ARG	HA	180	ARG	CA	-32,279	-5,521	-26,757
181	GLN	HA	181	GLN	CA	-16,381	-12,05	-4,3283
182	ALA	HA	182	ALA	CA	6,1136	-4,037	10,150
183	LEU	HA	183	LEU	CA	18,407	7,834	10,573
185	ASP	HA	185	ASP	CA	-20,761	-19,97	-0,7883
186	ILE	HA	186	ILE	CA	20,541	20,80	-0,2673
188	ALA	HA	188	ALA	CA	1,8105	9,493	-7,683
189	ILE	HA	189	ILE	CA	11,722	0,868	10,853
192	ASP	HA	192	ASP	CA	31,161	6,941	24,219
194	ASN	HA	194	ASN	CA	-16,449	-12,56	-3,8891
195	LYS	HA	195	LYS	CA	-7,9882	-7,663	-0,3251

196	VAL	HA	196	VAL	CA	-17,263	-5,560	-11,70
198	MET	HA	198	MET	CA	-43,701	-14,57	-29,12
200	SER	HA	200	SER	CA	-23,046	-10,08	-12,95
203	GLN	HA	203	GLN	CA	-16,414	-16,28	-0,129
205	SER	HA	205	SER	CA	-31,963	-9,665	-22,29
207	THR	HA	207	THR	CA	28,394	-1,039	29,43
208	GLN	HA	208	GLN	CA	23,814	1,496	22,31
209	PRO	HA	209	PRO	CA	-19,847	19,29	-39,14
210	LEU	HA	210	LEU	CA	-7,1023	-4,345	-2,757
211	TYR	HA	211	TYR	CA	-2,1171	1,617	-3,734
212	THR	HA	212	THR	CA	0,2113	-1,579	1,790
213	LEU	HA	213	LEU	CA	2,3124	-3,962	6,274
214	ASP	HA	214	ASP	CA	-25,011	-1,859	-23,15
216	ASP	HA	216	ASP	CA	25,797	-6,745	32,54
218	THR	HA	218	THR	CA	-7,5853	15,73	-23,31
219	LEU	HA	219	LEU	CA	-2,4869	-2,960	0,473
220	THR	HA	220	THR	CA	-6,5494	-3,097	-3,451
221	ASN	HA	221	ASN	CA	-13,196	2,347	-15,54
222	ASN	HA	222	ASN	CA	6,6947	14,59	-7,897
224	SER	HA	224	SER	CA	18,555	-8,261	26,81
226	VAL	HA	226	VAL	CA	-16,234	-4,181	-12,05
227	LYS	HA	227	LYS	CA	-10,907	-12,77	1,863
232	ASN	HA	232	ASN	CA	-7,2806	-7,633	0,352
234	ILE	HA	234	ILE	CA	-41,008	8,028	-49,03
236	PHE	HA	236	PHE	CA	-13,494	-11,10	-2,388
237	TYR	HA	237	TYR	CA	6,8712	-0,259	7,130
238	GLN	HA	238	GLN	CA	32,688	13,41	19,27
239	SER	HA	239	SER	CA	-5,7135	-6,649	0,935
240	ILE	HA	240	ILE	CA	-20,700	-8,623	-12,07
241	THR	HA	241	THR	CA	7,2655	-6,641	13,90
245	ASN	HA	245	ASN	CA	13,951	15,37	-1,418
246	TRP	HA	246	TRP	CA	-18,508	-13,85	-4,654
248	ASP	HA	248	ASP	CA	-11,915	-9,952	-1,963
249	GLU	HA	249	GLU	CA	18,143	21,78	-3,640
250	LYS	HA	250	LYS	CA	22,246	16,44	5,799
251	LEU	HA	251	LEU	CA	31,330	-4,546	35,87
252	SER	HA	252	SER	CA	-20,263	-17,03	-3,231
255	TYR	HA	255	TYR	CA	-15,291	-4,316	-10,97
256	THR	HA	256	THR	CA	9,7553	7,414	2,340
257	VAL	HA	257	VAL	CA	-10,873	5,554	-16,42
261	TRP	HA	261	TRP	CA	1,0435	6,869	-5,826
262	LYS	HA	262	LYS	CA	0,2962	-5,642	5,938
264	PHE	HA	264	PHE	CA	4,2701	13,42	-9,151
267	VAL	HA	267	VAL	CA	-1,3850	-8,174	6,789
268	PHE	HA	268	PHE	CA	6,9042	24,26	-17,36
269	THR	HA	269	THR	CA	7,6344	-22,63	30,27
271	GLU	HA	271	GLU	CA	0,1557	0,114	0,041
273	ILE	HA	273	ILE	CA	0,7576	-8,190	8,948
274	GLN	HA	274	GLN	CA	0,3038	-0,902	1,206

HαCα domain 1 (in absence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
99	GLN	HA	99	GLN	CA	-0,5500	-16,90	16,35
100	LEU	HA	100	LEU	CA	-1,0142	-5,526	4,511
101	THR	HA	101	THR	CA	-33,812	-18,68	-15,12
102	PHE	HA	102	PHE	CA	31,610	30,52	1,082
103	GLU	HA	103	GLU	CA	-24,517	-15,31	-9,202
104	ARG	HA	104	ARG	CA	-8,5401	-13,21	4,671
105	ALA	HA	105	ALA	CA	20,058	12,87	7,184
106	GLN	HA	106	GLN	CA	14,003	10,07	3,931
107	GLU	HA	107	GLU	CA	-12,843	-8,523	-4,319
108	VAL	HA	108	VAL	CA	-28,359	-16,31	-12,04
112	ARG	HA	112	ARG	CA	-28,895	-4,052	-24,84
113	SER	HA	113	SER	CA	3,5228	10,31	-6,787
209	PRO	HA	209	PRO	CA	-19,847	-14,66	-5,180
210	LEU	HA	210	LEU	CA	-7,1023	2,598	-9,700
211	TYR	HA	211	TYR	CA	-2,1171	-5,650	3,539
212	THR	HA	212	THR	CA	0,2113	-6,440	6,651
213	LEU	HA	213	LEU	CA	2,3124	4,025	-1,713
214	ASP	HA	214	ASP	CA	-25,011	-18,33	-6,678
216	ASP	HA	216	ASP	CA	25,797	22,94	2,853
218	THR	HA	218	THR	CA	-7,5853	-12,01	4,427
219	LEU	HA	219	LEU	CA	-2,4869	-3,068	0,581
220	THR	HA	220	THR	CA	-6,5494	-2,688	-3,860
221	ASN	HA	221	ASN	CA	-13,196	-10,65	-2,537
222	ASN	HA	222	ASN	CA	6,6947	-3,397	10,09
224	SER	HA	224	SER	CA	18,555	15,45	3,103
226	VAL	HA	226	VAL	CA	-16,234	-14,05	-2,179
227	LYS	HA	227	LYS	CA	-10,907	-10,64	-0,263
232	ASN	HA	232	ASN	CA	-7,2806	-10,96	3,687
234	ILE	HA	234	ILE	CA	-41,008	-17,67	-23,33
236	PHE	HA	236	PHE	CA	-13,494	-14,97	1,480
237	TYR	HA	237	TYR	CA	6,8712	20,95	-14,08
238	GLN	HA	238	GLN	CA	32,688	23,43	9,252
239	SER	HA	239	SER	CA	-5,7135	-10,42	4,714
240	ILE	HA	240	ILE	CA	-20,700	-16,69	-4,000
241	THR	HA	241	THR	CA	7,2655	-17,29	24,56
245	ASN	HA	245	ASN	CA	13,951	-2,477	16,42
246	TRP	HA	246	TRP	CA	-18,508	-12,71	-5,791
248	ASP	HA	248	ASP	CA	-11,915	-7,752	-4,163
249	GLU	HA	249	GLU	CA	18,143	12,41	5,727
250	LYS	HA	250	LYS	CA	22,246	21,29	0,952
251	LEU	HA	251	LEU	CA	31,330	29,46	1,865
252	SER	HA	252	SER	CA	-20,263	-8,150	-12,11
255	TYR	HA	255	TYR	CA	-15,291	-14,09	-1,193
256	THR	HA	256	THR	CA	9,7553	11,12	-1,372
257	VAL	HA	257	VAL	CA	-10,873	24,46	-35,33

HαCα domain 2 (in absence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
118	LYS	HA	118	LYS	CA	-3,6559	13,54	-17,19
119	THR	HA	119	THR	CA	-42,733	-25,51	-17,21
120	TYR	HA	120	TYR	CA	-48,703	-22,19	-26,50
121	ASN	HA	121	ASN	CA	10,949	10,32	0,629
127	ALA	HA	127	ALA	CA	-50,200	-29,05	-21,14
130	GLU	HA	130	GLU	CA	-22,655	-26,44	3,790
132	GLN	HA	132	GLN	CA	15,085	20,19	-5,105
133	SER	HA	133	SER	CA	52,614	42,39	10,21
134	ALA	HA	134	ALA	CA	39,265	41,22	-1,959
135	LEU	HA	135	LEU	CA	38,252	36,95	1,294
136	SER	HA	136	SER	CA	29,204	26,30	2,896
137	ASP	HA	137	ASP	CA	-3,7417	-26,08	22,34
139	GLU	HA	139	GLU	CA	23,911	-6,314	30,22
140	THR	HA	140	THR	CA	-29,459	-17,01	-12,44
142	LYS	HA	142	LYS	CA	-11,963	-14,69	2,735
143	ASN	HA	143	ASN	CA	14,327	8,937	5,389
144	TYR	HA	144	TYR	CA	27,688	22,97	4,710
146	SER	HA	146	SER	CA	32,880	23,72	9,159
147	ASP	HA	147	ASP	CA	-23,154	-30,52	7,371
148	ALA	HA	148	ALA	CA	41,522	37,56	3,961
149	PHE	HA	149	PHE	CA	22,204	29,58	-7,376
151	PHE	HA	151	PHE	CA	-21,752	-11,06	-10,68
154	GLU	HA	154	GLU	CA	10,189	12,39	-2,209
155	GLN	HA	155	GLN	CA	32,250	27,08	5,165
157	LEU	HA	157	LEU	CA	2,4039	28,65	-26,25
158	GLN	HA	158	GLN	CA	-21,686	-11,84	-9,842
159	LEU	HA	159	LEU	CA	-7,6420	-7,098	-0,543
161	GLU	HA	161	GLU	CA	46,384	34,51	11,87
162	THR	HA	162	THR	CA	45,347	31,79	13,55
163	THR	HA	163	THR	CA	-0,9199	5,061	-5,981
164	ALA	HA	164	ALA	CA	-14,192	-10,83	-3,362
166	PRO	HA	166	PRO	CA	31,852	43,61	-11,76
167	GLU	HA	167	GLU	CA	-7,6240	-11,56	3,941
169	GLU	HA	169	GLU	CA	33,201	-24,35	57,55
170	ALA	HA	170	ALA	CA	3,0521	25,02	-21,97
172	ASN	HA	172	ASN	CA	-2,2039	-18,87	16,66
173	LEU	HA	173	LEU	CA	-0,1915	18,91	-19,10
176	ILE	HA	176	ILE	CA	52,755	37,94	14,80
177	THR	HA	177	THR	CA	-6,4409	-9,669	3,228
178	GLN	HA	178	GLN	CA	-18,553	-4,848	-13,70
179	ASN	HA	179	ASN	CA	-41,422	-22,53	-18,88
180	ARG	HA	180	ARG	CA	-32,279	-16,31	-15,96
181	GLN	HA	181	GLN	CA	-16,381	-11,30	-5,074
182	ALA	HA	182	ALA	CA	6,1136	-6,216	12,32
183	LEU	HA	183	LEU	CA	18,407	22,95	-4,548
185	ASP	HA	185	ASP	CA	-20,761	-18,80	-1,954
186	ILE	HA	186	ILE	CA	20,541	26,91	-6,375
188	ALA	HA	188	ALA	CA	1,8105	9,852	-8,041
189	ILE	HA	189	ILE	CA	11,722	1,760	9,962
192	ASP	HA	192	ASP	CA	31,161	19,57	11,58

194	ASN	HA	194	ASN	CA	-16,449	-13,88	-2,562
195	LYS	HA	195	LYS	CA	-7,9882	-7,019	-0,969
196	VAL	HA	196	VAL	CA	-17,263	-4,120	-13,14
198	MET	HA	198	MET	CA	-43,701	-32,57	-11,12
200	SER	HA	200	SER	CA	-23,046	-12,58	-10,46
203	GLN	HA	203	GLN	CA	-16,414	-17,26	0,847
205	SER	HA	205	SER	CA	-31,963	-15,52	-16,44
207	THR	HA	207	THR	CA	28,394	-11,26	39,65

H^NN all (in presence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
101	THR	H	101	THR	N	14,425	7,851	6,574
102	PHE	H	102	PHE	N	9,0571	5,234	3,822
103	GLU	H	103	GLU	N	2,6068	-1,117	3,724
107	GLU	H	107	GLU	N	3,0227	1,405	1,617
113	SER	H	113	SER	N	-10,636	-7,128	-3,508
116	ALA	H	116	ALA	N	-4,3288	-2,676	-1,652
117	GLY	H	117	GLY	N	-2,2200	-7,919	5,699
123	GLY	H	123	GLY	N	-7,6853	-7,697	0,012
125	TYR	H	125	TYR	N	-15,312	-0,011	-15,30
127	ALA	H	127	ALA	N	3,5298	7,062	-3,533
130	GLU	H	130	GLU	N	10,731	8,438	2,293
132	GLN	H	132	GLN	N	4,1920	-7,160	11,35
134	ALA	H	134	ALA	N	-16,056	-5,328	-10,72
135	LEU	H	135	LEU	N	-3,8126	-1,305	-2,507
136	SER	H	136	SER	N	-4,4528	-1,989	-2,463
137	ASP	H	137	ASP	N	1,5724	1,345	0,226
138	GLY	H	138	GLY	N	7,3332	6,959	0,373
140	THR	H	140	THR	N	-1,1438	-2,325	1,181
141	GLY	H	141	GLY	N	11,521	6,878	4,643
146	SER	H	146	SER	N	-4,4620	-7,990	3,528
147	ASP	H	147	ASP	N	14,219	7,635	6,584
148	ALA	H	148	ALA	N	-10,604	4,042	-14,64
152	GLY	H	152	GLY	N	-5,0220	-3,170	-1,851
159	LEU	H	159	LEU	N	5,7900	0,260	-4,480
161	GLU	H	161	GLU	N	-12,431	-6,185	-6,246
164	ALA	H	164	ALA	N	1,9592	-1,987	3,946
168	GLY	H	168	GLY	N	9,5131	8,013	1,499
170	ARG	H	170	ARG	N	-0,6202	-1,773	1,152
171	ALA	H	171	ALA	N	-7,5485	-2,344	-5,203
177	THR	H	177	THR	N	29,489	6,930	22,55
182	ALA	H	182	ALA	N	-2,3532	0,228	-2,581
188	ALA	H	188	ALA	N	4,7739	-5,793	10,56
190	LEU	H	190	LEU	N	12,291	12,31	-0,028
193	GLY	H	193	GLY	N	21,441	-5,088	26,53
196	VAL	H	196	VAL	N	9,6463	0,671	8,975
198	MET	H	198	MET	N	-4,2704	-5,519	1,249
200	SER	H	200	SER	N	10,321	2,001	8,320
203	GLN	H	203	GLN	N	9,6773	2,377	7,299
204	PHE	H	204	PHE	N	-3,8162	4,140	-7,956
205	SER	H	205	SER	N	1,4812	-4,042	5,523
206	GLY	H	206	GLY	N	3,8089	-5,156	8,965

210	LEU	H	210	LEU	N	-12,167	-8,473	-3,694
211	TYR	H	211	TYR	N	-1,5214	-0,361	-1,160
213	LEU	H	213	LEU	N	10,060	3,808	6,252
214	ASP	H	214	ASP	N	0,7406	0,630	0,110
215	GLY	H	215	GLY	N	1,2003	-2,988	4,188
217	GLY	H	217	GLY	N	15,160	6,886	8,274
219	LEU	H	219	LEU	N	9,9893	3,974	6,014
220	THR	H	220	THR	N	4,8724	5,411	-0,539
221	ASN	H	221	ASN	N	16,014	9,236	6,778
222	ASN	H	222	ASN	N	-9,6992	-7,407	-2,291
225	GLY	H	225	GLY	N	1,6436	-0,972	2,615
226	VAL	H	226	VAL	N	-1,3864	3,897	-5,284
228	TYR	H	228	TYR	N	10,972	2,506	8,466
234	ILE	H	234	ILE	N	4,1099	-0,090	4,200
238	GLN	H	238	GLN	N	-9,3636	-1,332	-8,030
240	ILE	H	240	ILE	N	18,313	13,34	4,973
241	THR	H	241	THR	N	14,281	10,43	3,848
245	ASN	H	245	ASN	N	-4,2339	-5,160	0,926
247	GLY	H	247	GLY	N	2,1489	-7,293	9,441
248	ASP	H	248	ASP	N	-3,2069	-3,363	0,156
250	LYS	H	250	LYS	N	1,2459	-5,094	6,340
254	GLY	H	254	GLY	N	-6,6783	0,360	-7,039
256	THR	H	256	THR	N	13,535	8,643	4,891
260	GLY	H	260	GLY	N	-5,6477	4,725	-10,37
262	LYS	H	262	LYS	N	-0,5546	-2,561	2,007
264	PHE	H	264	PHE	N	-7,7820	-6,719	-1,062
268	PHE	H	268	PHE	N	-3,0227	-3,721	0,698
272	GLY	H	272	GLY	N	3,2306	-3,945	7,175
273	ILE	H	273	ILE	N	-1,0671	-7,046	5,978
274	GLN	H	274	GLN	N	-2,8184	-8,404	5,586
275	LYS	H	275	LYS	N	-4,2029	2,069	-6,272

H^NN domain 1 (in presence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
101	THR	H	101	THR	N	14,425	10,35	4,070
102	PHE	H	102	PHE	N	9,0571	6,653	2,403
103	GLU	H	103	GLU	N	2,6068	1,792	0,814
107	GLU	H	107	GLU	N	3,0227	4,592	-1,569
113	SER	H	113	SER	N	-10,636	-13,67	3,040
210	LEU	H	210	LEU	N	-12,167	-10,39	-1,774
211	TYR	H	211	TYR	N	-1,5214	0,927	-2,448
213	LEU	H	213	LEU	N	10,060	6,186	3,873
214	ASP	H	214	ASP	N	0,7406	5,117	-4,376
215	GLY	H	215	GLY	N	1,2003	-1,781	2,981
217	GLY	H	217	GLY	N	15,160	10,74	4,413
219	LEU	H	219	LEU	N	9,9893	7,408	2,581
220	THR	H	220	THR	N	4,8724	9,876	-5,004
221	ASN	H	221	ASN	N	16,014	13,57	2,434
222	ASN	H	222	ASN	N	-9,6992	-11,23	1,533
225	GLY	H	225	GLY	N	1,6436	0,186	1,457
226	VAL	H	226	VAL	N	-1,3864	-3,049	1,663
228	TYR	H	228	TYR	N	10,972	7,036	3,936

234	ILE	H	234	ILE	N	4,1099	1,110	2,990
238	GLN	H	238	GLN	N	-9,3636	-2,979	-6,383
240	ILE	H	240	ILE	N	18,313	17,75	0,558
241	THR	H	241	THR	N	14,281	14,37	-0,095
245	ASN	H	245	ASN	N	-4,2339	-6,558	2,324
247	GLY	H	247	GLY	N	2,1489	-4,246	6,395
248	ASP	H	248	ASP	N	-3,2069	-1,034	-2,172
250	LYS	H	250	LYS	N	1,2459	-7,124	8,370
254	GLY	H	254	GLY	N	-6,6783	1,894	-8,573
256	THR	H	256	THR	N	13,535	13,07	0,461

H^NN domain 2 (in presence of MalE)

ID_I	RES_I	ATOM_I	ID_J	RES_J	ATOM_J	D_OBS	D	D_DIFF
117	GLY	H	117	GLY	N	-2,2200	-10,53	8,315
123	GLY	H	123	GLY	N	-7,6853	-5,527	-2,157
125	TYR	H	125	TYR	N	15,312	3,771	-19,08
127	ALA	H	127	ALA	N	3,5298	7,832	-4,302
130	GLU	H	130	GLU	N	10,731	6,792	3,938
132	GLN	H	132	GLN	N	4,1920	-3,804	7,996
134	ALA	H	134	ALA	N	16,056	-6,087	-9,969
135	LEU	H	135	LEU	N	-3,8126	-2,388	-1,424
136	SER	H	136	SER	N	-4,4528	-3,736	-0,716
137	ASP	H	137	ASP	N	1,5724	0,642	0,930
138	GLY	H	138	GLY	N	7,3332	7,500	-0,167
140	THR	H	140	THR	N	-1,1438	-3,640	2,496
141	GLY	H	141	GLY	N	11,521	4,730	6,791
146	SER	H	146	SER	N	-4,4620	-6,768	2,306
147	ASP	H	147	ASP	N	14,219	8,323	5,896
148	ALA	H	148	ALA	N	10,604	2,365	-12,96
152	GLY	H	152	GLY	N	-5,0220	-2,438	-2,583
159	LEU	H	159	LEU	N	5,7900	7,412	-1,622
161	GLU	H	161	GLU	N	12,431	-4,239	-8,192
164	ALA	H	164	ALA	N	1,9592	3,411	-1,452
168	GLY	H	168	GLY	N	9,5131	9,827	-0,314
170	ARG	H	170	ARG	N	-0,6202	0,352	-0,972
171	ALA	H	171	ALA	N	-7,5485	-1,684	-5,864
177	THR	H	177	THR	N	29,489	8,956	20,53
182	ALA	H	182	ALA	N	-2,3532	-0,620	-1,733
188	ALA	H	188	ALA	N	4,7739	-3,444	8,218
190	LEU	H	190	LEU	N	12,291	9,420	2,871
193	GLY	H	193	GLY	N	21,441	-2,479	23,92
196	VAL	H	196	VAL	N	9,6463	-0,047	9,694
198	MET	H	198	MET	N	-4,2704	-2,852	-1,417
200	SER	H	200	SER	N	10,321	0,316	10,00
203	GLN	H	203	GLN	N	9,6773	0,589	9,087
204	PHE	H	204	PHE	N	-3,8162	3,899	-7,715
205	SER	H	205	SER	N	1,4812	-3,860	5,342
206	GLY	H	206	GLY	N	3,8089	-3,275	7,084

A.5 References

1. Ernst, R. R. Citation Classic - Application of Fourier-Transform Spectroscopy to Magnetic-Resonance. Current Contents/Physical Chemical & Earth Sciences, 24-24 (1983).
2. Jeener, J. (1971).
3. Bloch, F., Hansen, W. W. & Packard, M. Nuclear Induction. Physical Review 69, 127-127 (1946).
4. Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. Physical Review 69, 37-38 (1946).
5. Sloan, T. The spin structure of the nucleon. Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences 359, 379-389 (2001).
6. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551-3567 (1999).
7. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. Chemical Reviews 104, 3607-3622 (2004).
8. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology 6, 197-208 (2005).
9. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. Nmr Determination of Residual Structure in a Urea-Denatured Protein, the 434-Repressor. Science 257, 1559-1563 (1992).
10. Rith, K. & Schafer, A. The mystery of nucleon spin. Scientific American 281, 58-63 (1999).
11. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Skelton, N. J. & Rance, M. Protein NMR Spectroscopy (Academic Press, San Diego, 2006).
12. Grant, D. M. & Harris, R. Encyclopedia of Nuclear Magnetic Resonance, Advances in NMR (Wiley & Sons, 2002).
13. Teng, Q. Structural Biology: Practical NMR Applications (Springer, 2005).
14. Wuthrich, K. NMR of Proteins and Nucleic Acids (Wiley-Interscience, San Diego, 1986).
15. Merzbacher, E. Quantum mechanics (Wiley & Sons, New York, 1970).
16. Ramsey, N. F. & Purcell, E. M. Interactions between Nuclear Spins in Molecules. Physical Review 85, 143-144 (1952).
17. Ewing, B., Glaser, S. J. & Drobny, G. P. Experimental Demonstrations of Shaped Pulses for Narrow-Band Inversion of Uncoupled and Coupled Systems. Journal of Magnetic Resonance 98, 381-387 (1992).
18. Iwamiya, J. H., Sinton, S. W., Liu, H., Glaser, S. J. & Drobny, G. P. Multiple-Pulse Sequences for Homonuclear Decoupling - Experimental-Verification. Journal of Magnetic Resonance 100, 367-375 (1992).
19. LIU, H., Glaser, S. J. & Drobny, G. P. Development and Optimization of Multipulse Propagators - Applications to Homonuclear Spin Decoupling in Solids. Journal of Chemical Physics 93, 7543-7560 (1990).

20. Clore, G. M. & Gronenborn, A. M. New methods of structure refinement for macromolecular structure determination by NMR. Proceedings of the National Academy of Sciences of the United States of America 95, 5891-5898 (1998).
21. Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. Essential Dynamics of Proteins. Proteins-Structure Function and Genetics 17, 412-425 (1993).
22. Fischer, M. W. F., Majumdar, A. & Zuiderweg, E. R. P. Protein NMR relaxation: theory, applications and outlook. Progress in Nuclear Magnetic Resonance Spectroscopy 33, 207-272 (1998).
23. Palmer, A. G., Kroenke, C. D. & Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B 339, 204-238 (2001).
24. Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. Uniform and residue-specific N-15-labeling of proteins on a highly deuterated background. Journal of Biomolecular Nmr 29, 289-297 (2004).
25. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. & Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated N-15-, C-13-, H-2-labeled proteins. Journal of Biomolecular Nmr 13, 369-374 (1999).
26. Kainosho, M. et al. Optimal isotope labelling for NMR protein structure determinations. Nature 440, 52-57 (2006).
27. Mathys, S. et al. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene 231, 1-13 (1999).
28. Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. & Kyogoku, Y. Improved segmental isotope labeling of proteins and application to a larger protein. Journal of Biomolecular Nmr 14, 105-114 (1999).
29. Gryk, M. R., Abseher, R., Simon, B., Nilges, M. & Oschkinat, H. Heteronuclear relaxation study of the PH domain of beta-spectrin: Restriction of loop motions upon binding inositol trisphosphate. Journal of Molecular Biology 280, 879-896 (1998).
30. Wider, G. & Wuthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. Current Opinion in Structural Biology 9, 594-601 (1999).
31. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proceedings of the National Academy of Sciences of the United States of America 94, 12366-12371 (1997).
32. Wuthrich, K., Pervushin, K., Riek, R., Salzmann, M. & Wider, G. New perspectives for NMR in structural biology using TROSY and CRINEPT. Abstracts of Papers of the American Chemical Society 219, U269-U269 (2000).

33. Gardner, K. H., Rosen, M. K. & Kay, L. E. Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. Biochemistry 36, 1389-1401 (1997).
34. Markley, J. L., Meadows, D. H. & Jardetzk.O. Nuclear Magnetic Resonance Studies of Helix-Coil Transitions in Polyamino Acids. Journal of Molecular Biology 27, 25-& (1967).
35. Wishart, D. S. & Sykes, B. D. Chemical-Shifts as a Tool for Structure Determination. Nuclear Magnetic Resonance, Pt C 239, 363-392 (1994).
36. Wishart, D. S., Sykes, B. D. & Richards, F. M. The Chemical-Shift Index - a Fast and Simple Method for the Assignment of Protein Secondary Structure through Nmr-Spectroscopy. Biochemistry 31, 1647-1651 (1992).
37. Wishart, D. S., Sykes, B. D. & Richards, F. M. Relationship between Nuclear-Magnetic-Resonance Chemical-Shift and Protein Secondary Structure. Journal of Molecular Biology 222, 311-333 (1991).
38. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of Biomolecular Nmr 13, 289-302 (1999).
39. Phizicky, E. M. & Fields, S. Protein-Protein Interactions - Methods for Detection and Analysis. Microbiological Reviews 59, 94-123 (1995).
40. RIKEN.
41. (M-KBSGI).
42. CESG.
43. Pascal, S. M. & McKelvie, J. M. NMR Primer: An HSQC based approach (IM Publications LLP, Toronto, 2008).
44. Takeuchi, K. & Wagner, G. NMR studies of protein interactions. Current Opinion in Structural Biology 16, 109-117 (2006).
45. Fernandez, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. Current Opinion in Structural Biology 13, 570-580 (2003).
46. Markus, M. A., Dayie, K. T., Matsudaira, P. & Wagner, G. Effect of Deuteration on the Amide Proton Relaxation Rates in Proteins - Heteronuclear Nmr Experiments on Villin 14t. Journal of Magnetic Resonance Series B 105, 192-195 (1994).
47. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445, 618-622 (2007).
48. Holde, K. E., Johnson, C. & Ho, P. S. Principles of Physical Biochemistry (Prentice Hall, 2005).
49. Mayer, M. & Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angewandte Chemie-International Edition 38, 1784-1788 (1999).
50. Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. & Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. Journal of Biomolecular Nmr 21, 349-359 (2001).
51. Dalvit, C. et al. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. Journal of Biomolecular Nmr 18, 65-68 (2000).

52. Chen, A. & Shapiro, M. J. NOE pumping: A novel NMR technique for identification of compounds with binding affinity to macromolecules. Journal of the American Chemical Society 120, 10258-10259 (1998).
53. Meyer, B., Weimar, T. & Peters, T. Screening mixtures for biological activity by NMR. European Journal of Biochemistry 246, 705-709 (1997).
54. Clore, G. M., Gronenborn, A. M. & Bax, A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. Journal of Magnetic Resonance 133, 216-221 (1998).
55. Tjandra, N. Establishing a degree of order: obtaining high-resolution NMR structures from molecular alignment. Structure with Folding & Design 7, R205-R211 (1999).
56. Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution. Proceedings of the National Academy of Sciences of the United States of America 92, 9279-9283 (1995).
57. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium (vol 278, pg 1111, 1997). Science 278, 1697-1697 (1997).
58. Cai, M. et al. Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. Nature Structural Biology 5, 903-909 (1998).
59. Mollova, E. T., Hansen, M. R. & Pardi, A. Global structure of RNA determined with residual dipolar couplings. Journal of the American Chemical Society 122, 11561-11562 (2000).
60. Skrynnikov, N. R. et al. Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: Differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. Journal of Molecular Biology 295, 1265-1273 (2000).
61. Olejniczak, E. T. et al. Improved NMR structures of protein/ligand complexes using residual dipolar couplings. Journal of the American Chemical Society 121, 9249-9250 (1999).
62. Weaver, J. L. & Prestegard, J. H. Nuclear magnetic resonance structural and ligand binding studies of BLBC, a two-domain fragment of barley lectin. Biochemistry 37, 116-128 (1998).
63. Gayathri, C., Bothnerby, A. A., Vanzijl, P. C. M. & Maclean, C. Dipolar Magnetic-Field Effects in Nmr-Spectra of Liquids. Chemical Physics Letters 87, 192-196 (1982).
64. Bax, A., Kontaxis, G. & Tjandra, N. Dipolar couplings in macromolecular structure determination. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B 339, 127-174 (2001).
65. Prestegard, J. H., Al-Hashimi, H. M. & Tolman, J. R. NMR structures of biomolecules using field oriented media and residual dipolar couplings. Quarterly Reviews of Biophysics 33, 371-424 (2000).
66. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. Use of dipolar H-1-N-15 and H-1-C-13 couplings in the structure

determination of magnetically oriented macromolecules in solution. Nature Structural Biology 4, 732-738 (1997).

67. Losonczi, J. A., Andrec, M., Fischer, M. W. F. & Prestegard, J. H. Order matrix analysis of residual dipolar couplings using singular value decomposition. Journal of Magnetic Resonance 138, 334-342 (1999).
68. Fischer, M. W. F., Losonczi, J. A., Weaver, J. L. & Prestegard, J. H. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. Biochemistry 38, 9013-9022 (1999).
69. Ojennus, D. D., Mitton-Fry, R. M. & Wuttke, D. S. Induced alignment and measurement of dipolar couplings of an SH2 domain through direct binding with filamentous phage. Journal of Biomolecular Nmr 14, 175-179 (1999).
70. Higman, V. (University of Oxford, 2000).
71. Losonczi, J. A. & Prestegard, J. H. Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. Journal of Biomolecular Nmr 12, 447-451 (1998).
72. Almond, A. & Axelsen, J. B. Physical interpretation of residual dipolar couplings in neutral aligned media. Journal of the American Chemical Society 124, 9986-9987 (2002).
73. Azurmendi, H. F. & Bush, C. A. Tracking alignment from the moment of inertia tensor (TRAMITE) of biomolecules in neutral dilute liquid crystal solutions. Journal of the American Chemical Society 124, 2426-2427 (2002).
74. Fernandes, M. X., Bernado, P., Pons, M. & de la Torre, J. G. An analytical solution to the problem of the orientation of rigid particles by planar obstacles. Application to membrane systems and to the calculation of dipolar couplings in protein NMR spectroscopy. Journal of the American Chemical Society 123, 12037-12047 (2001).
75. Zweckstetter, M. & Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. Journal of the American Chemical Society 122, 3791-3792 (2000).
76. Ramirez, B. E. & Bax, A. Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. Journal of the American Chemical Society 120, 9106-9107 (1998).
77. Tjandra, N., Grzesiek, S. & Bax, A. Magnetic field dependence of nitrogen-proton J splittings in N-15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. Journal of the American Chemical Society 118, 6264-6272 (1996).
78. Clore, G. M., Starich, M. R. & Gronenborn, A. M. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. Journal of the American Chemical Society 120, 10571-10572 (1998).
79. Skrynnikov, N. R. & Kay, L. E. Assessment of molecular structure using frame-independent orientational restraints derived from residual dipolar couplings. Journal of Biomolecular Nmr 18, 239-252 (2000).
80. Warren, J. J. & Moore, P. B. A maximum likelihood method for determining D-a(PQ) and R for sets of dipolar coupling data. Journal of Magnetic Resonance 149, 271-275 (2001).

81. Fushman, D., Ghose, R. & Cowburn, D. The effect of finite sampling on the determination of orientational properties: A theoretical treatment with application to interatomic vectors in proteins. Journal of the American Chemical Society 122, 10640-10649 (2000).
82. Nomura, K. & Kainosho, M. Graphical analysis of the relative orientation of molecular alignment tensors for a protein dissolved in two different anisotropic media. Journal of Magnetic Resonance 154, 146-153 (2002).
83. Hus, J. C. & Bruschweiler, R. Reconstruction of interatomic vectors by principle component analysis of nuclear magnetic resonance data in multiple alignments. Journal of Chemical Physics 117, 1166-1172 (2002).
84. Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nature Structural Biology 5, 1065-1074 (1998).
85. Zweckstetter, M. & Bax, A. Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. Journal of Biomolecular Nmr 20, 365-377 (2001).
86. Beraud, S. et al. Direct structure determination using residual dipolar couplings: Reaction-site conformation of methionine sulfoxide reductase in solution. Journal of the American Chemical Society 124, 13709-13715 (2002).
87. Chou, J. J., Li, S. P. & Bax, A. Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings. Journal of Biomolecular Nmr 18, 217-227 (2000).
88. Medek, A., Olejniczak, E. T., Meadows, R. P. & Fesik, S. W. An approach for high-throughput structure determination of proteins by NMR spectroscopy. Journal of Biomolecular Nmr 18, 229-238 (2000).
89. Mueller, G. A. et al. Global folds of proteins with low densities of NOEs using residual dipolar couplings: Application to the 370-residue maltodextrin-binding protein. Journal of Molecular Biology 300, 197-212 (2000).
90. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. An easy way to include weak alignment constraints into NMR structure calculations. Journal of Biomolecular Nmr 21, 275-280 (2001).
91. Tian, F., Valafar, H. & Prestegard, J. H. A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. Journal of the American Chemical Society 123, 11791-11796 (2001).
92. Markus, M. A., Gerstner, R. B., Draper, D. E. & Torchia, D. A. Refining the overall structure and subdomain orientation of ribosomal protein S4 Delta 41 with dipolar couplings measured by NMR in uniaxial liquid crystalline phases. Journal of Molecular Biology 292, 375-387 (1999).
93. Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. Journal of Biomolecular Nmr 20, 223-231 (2001).
94. Goto, N. K., Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. Journal of Molecular Biology 308, 745-764 (2001).

95. Bewley, C. A. & Clore, G. M. Determination of the relative orientation of the two halves of the domain-swapped dimer of cyanovirin-N in solution using dipolar couplings and rigid body minimization. Journal of the American Chemical Society 122, 6009-6016 (2000).
96. Clore, G. M. & Bewley, C. A. Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. Journal of Magnetic Resonance 154, 329-335 (2002).
97. Biekofsky, R. R. et al. NMR approaches for monitoring domain orientations in calcium-binding proteins in solution using partial replacement of Ca2+ by Tb3+. Febs Letters 460, 519-526 (1999).
98. Braddock, D. T., Cai, M. L., Baber, J. L., Huang, Y. & Clore, G. M. Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. Journal of the American Chemical Society 123, 8634-8635 (2001).
99. Bolon, P. J., Al-Hashimi, H. M. & Prestegard, J. H. Residual dipolar coupling derived orientational constraints on ligand geometry in a 53 kDa protein-ligand complex. Journal of Molecular Biology 293, 107-115 (1999).
100. Clore, G. M. Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear Overhauser enhancement data and dipolar couplings by rigid body minimization. Proceedings of the National Academy of Sciences of the United States of America 97, 9021-9025 (2000).
101. Koenig, B. W. et al. Measurement of dipolar couplings in a transducin peptide fragment weakly bound to oriented photo-activated rhodopsin. Journal of Biomolecular Nmr 16, 121-125 (2000).
102. Levitt, M. H. Spin Dynamics: Basics of Nuclear Magnetic Resonance (Wiley, 2008).
103. Fushman, D., Weisemann, R., Thuring, H. & Ruterjans, H. Backbone Dynamics of Ribonuclease-T1 and Its Complex with 2'gmp Studied by 2-Dimensional Heteronuclear Nmr-Spectroscopy. Journal of Biomolecular Nmr 4, 61-78 (1994).
104. Allerhan.A, Chen, F. M. & Gutowsky, H. S. Spin-Echo Nmr Studies of Chemical Exchange .3. Conformational Isomerization of Cyclohexane and D11-Cyclohexane. Journal of Chemical Physics 42, 3040-& (1965).
105. Loria, J. P., Rance, M. & Palmer, A. G. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. Journal of the American Chemical Society 121, 2331-2332 (1999).
106. Martin, M. L., Martin, G. J. & Delpeuch, J. J. Practical NMR Spectroscopy (AAA, 1980).
107. KAY, L. E., Torchia, D. A. & Bax, A. Backbone Dynamics of Proteins as Studied by N-15 Inverse Detected Heteronuclear Nmr-Spectroscopy - Application to Staphylococcal Nuclease. Biochemistry 28, 8972-8979 (1989).
108. Mittermaier, A. & Kay, L. E. Review - New tools provide new insights in NMR studies of protein dynamics. Science 312, 224-228 (2006).

109. Palmer, A. G. NMR probes of molecular dynamics: Overview and comparison with other techniques. Annual Review of Biophysics and Biomolecular Structure 30, 129-155 (2001).
110. Vallurupalli, P., Hansen, D. F. & Kay, L. E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. Proceedings of the National Academy of Sciences of the United States of America 105, 11766-11771 (2008).
111. Carver, J. P. & Richards, R. E. General 2-Site Solution for Chemical Exchange Produced Dependence of T2 Upon Carr-Purcell Pulse Separation. Journal of Magnetic Resonance 6, 89-& (1972).
112. Millet, O., Loria, J. P., Kroenke, C. D., Pons, M. & Palmer, A. G. The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. Journal of the American Chemical Society 122, 2867-2877 (2000).
113. Hartmann, S. R. & Hahn, E. L. Nuclear Double Resonance in Rotating Frame. Physical Review 128, 2042-& (1962).
114. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. Journal of Magnetic Resonance 131, 373-378 (1998).
115. Meissner, A., Duus, J. O. & Sorensen, O. W. Spin-state-selective excitation. Application for E.COSY-type measurement of J(HH) coupling constants. Journal of Magnetic Resonance 128, 92-97 (1997).
116. Strauss, H. M. Struktur Funktions Untersuchungen an cyanobakteriellem Phytochrom 1 (Freie Universitat, Berlin, 2006).
117. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. Science 295, 1520-1523 (2002).
118. Grzesiek, S. Notes on relaxation and dynamics. (Basel 2005).

Acknowledgements

This work was carried out at the Department of NMR Supported Structural Biology, at the Leibniz-Institut für Molekulare Pharmakologie (FMP), Berlin, Germany, from 1st May 2005 to 20th May 2009 under the supervision of **Prof. Dr. Bernd Reif**. This research work was supported by the Leibniz- Gemeinschaft and the DFG (grants RE1435, SFB 449, SFB 740). This thesis would not have been possible without the help of many people. I sincerely acknowledge them.

First I would like to express my deepest gratitude to my guide **Prof. Dr. Bernd Reif** who gave me the opportunity to work at the FMP. He has with great patience encouraged and guided me through my work. His scientific discussions and motivation were the essential ingredients for my research work. Under his shelter I have been able to explore my scientific interests and his trust has fuelled my confidence throughout my work.

Next off I would like to thank **Prof. Dr. Erwin Schneider** for the fruitful collaboration with his students **Mathias Grote** and **Dr. Martin Daus**. They have been of great support and their discussions have led to many new ideas during the time. I am in debt to **Dr. Peter Schmieder** for all his help regarding NMR experimental procedures and to be able to answer all the questions concerning my ideas. He has been off great support. I would like to thank **Dr. Alexei Chernagolov** to introducing me into the world of recombinant protein work. I would also like to thank **Dr. Christian Freund** and his group for fruitful discussions regarding protein chemistry. **Dr. Victoria Higman** for putting up with me learning CCPNMR and all the different tidbits of measuring and analyzing RDCs of protein. She has been very patient and understanding. I am grateful to **Prof. Dr. Hartmut Oschkinat** for the NMR-facilities at FMP, Berlin, and to the people of his group for their kindness and help. To the people in the group of Solid-State NMR at the FMP. **Vipin, Murali, Mangesh** and **Dr. Veniamin Chevelkov** for their companionship during my whole work. Juan-Miguel for many laughs not involving science and the people in my office **Rasmus, Andy** and **Sami** for putting up with all my mischiefs in the office. It□s been a wonderful time with many memories and laughs.

I would like to thank all of my friends outside the FMP, **Kicki, Johanna, Niklas** and more for all the good times in the few hours I have spent outside work. And a special thank to **Hanna** for putting up with me during these last few months of stress.

Last I would like to thank my family and friends in Stockholm for supporting me through this time. Without you this would never have happened.

Publications

Jacso T., Grote M., Schmieder P., Schneider E., Reif B. NMR assignments of the periplasmic loop P2 of the MalF subunit of the maltose ATP binding cassette transporter. Biomol. NMR assign. DOI: 10.1007/s12104-008-9131- 7, 2008.

Jacso T., Grote M., Daus M., Schmieder P., Keller S., Schneider E., Reif B. The periplasmic loop P2 of the MalF subunit of the maltose ATP binding cassette transporter is sufficient to bind the maltose binding protein MalE. Biochemistry, 2009, 48 (10), pp 2216–2225.

Thoms S., Max K., Wunderlich M., **Jacso T.**, Reif B., Heinemann U., Schmid F. X. Dimer formation of a stabilized Gβ1 variant. A structural and energetic analysis. JMB (Submitted)

Jacso T., Max K., Heinemann U., Reif B. NMR assignments and RDC studies of a hyperstable Gβ1 variant. Biomol. NMR assign. (Submitted)

Conferences and Workshops

Jacso, T., Chernagolov, A., Gast, K., Reif, B. Understanding molecular mechanisms of fibril formation in the PI3K-SH3 domain. Poster presentation, International Conferences on Magnetic Resonance in Biological Systems (ICMRBS), Göttingen, Deutschland (2006).

Jacso, T., Grote, M., Schneider, E., Reif, B. A functional and structural study of the loop regions of the maltose-ABC-transporter of *Enterobacteriae*. Poster presentation, Annual Meeting of Magnetic Resonance in Biophysical Chemistry (GDCH-FGMR), Göttingen, Deutschland (2007).

Jacso, T., Grote, M., Schmieder, P., Keller, S., Schneider, E., Reif, B. NMR investigations of the isolated loop P2 of the MalF subunit of the maltose transporter ($MalFGK_2$). Poster presentation, International Conferences on Magnetic Resonance in Biological Systems (ICMRBS), Sand Diego, USA (2008).

Jacso, T., Grote, M., Schmieder, P., Keller, S., Schneider, E., Reif, B. NMR investigations of the isolated loop P2 of the MalF subunit of the maltose transporter ($MalFGK_2$). Poster presentation, EMBO course "Multidimensional NMR in structural biology", Il Ciocco, Italy (2008).

Eidesstattliche Erklärung

Hiermit versichere ich, daß ich die vorliegende Dissertation eingeständig angefertigt und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Berlin, …….

Tomas Jacso